

INVESTIGACIÓN X PROFUNDIZACIÓN \_\_\_\_

Modelo de evaluación de la eficiencia de los paneles solares, basado en variables climatológicas, calidad del aire e irradiancia

**Autor**

**Luis Carrillo Mejía**

**Director**

**Dr. Elvis Eduardo Gaona**

**Codirector**

**Dr Johan Alexander Hernández Mora**

**Universidad Distrital Francisco José De Caldas**

**Maestría en Ciencias de la Información y las Comunicaciones**

**Énfasis en Sistemas de Información**

**Bogotá, Colombia**

**abril de 2024**

Contenido

[RESUMEN 4](#_Toc165152510)

[PALABRAS CLAVE 4](#_Toc165152511)

[INTRODUCCIÓN 5](#_Toc165152512)

[1. PROBLEMA DE INVESTIGACIÓN 6](#_Toc165152513)

[1.1 PLANTEAMIENTO DEL PROBLEMA 6](#_Toc165152514)

[1.2 FORMULACIÓN DEL PROBLEMA 7](#_Toc165152515)

[1.3 SISTEMATIZACIÓN DEL PROBLEMA 7](#_Toc165152516)

[2. OBJETIVOS 8](#_Toc165152517)

[2.1 OBJETIVO GENERAL 8](#_Toc165152518)

[2.2 OBJETIVOS ESPECÍFICOS 8](#_Toc165152519)

[3. JUSTIFICACIÓN 9](#_Toc165152520)

[3.1 JUSTIFICACIÓN TEÓRICA 9](#_Toc165152521)

[3.2 JUSTIFICACIÓN METODOLÓGICA 9](#_Toc165152522)

[3.3 JUSTIFICACIÓN PRÁCTICA 10](#_Toc165152523)

[4. MARCO DE REFERENCIA 11](#_Toc165152524)

[4.1 MARCO CONCEPTUAL 11](#_Toc165152525)

[4.1.1 PANELES SOLARES 11](#_Toc165152526)

[4.1.2 IRRADIANCIA 15](#_Toc165152527)

[4.1.3 CALIDAD DEL AIRE (Sagheer et al., 2024) 17](#_Toc165152528)

[4.1.4 ALGORITMOS DE PREDICCIÓN 18](#_Toc165152529)

[4.2 MARCO TEÓRICO 20](#_Toc165152530)

[4.3 MARCO ESPACIAL 22](#_Toc165152531)

[4.4 MARCO TEMPORAL 23](#_Toc165152532)

[5. HIPÓTESIS 24](#_Toc165152533)

[6. METODOLOGÍA 25](#_Toc165152534)

[7. IMPACTO Y RESULTADOS ESPERADOS 28](#_Toc165152535)

[8. RECURSOS Y PRESUPUESTO 29](#_Toc165152536)

[9. CRONOGRAMA 30](#_Toc165152537)

[10. LIMITACION 31](#_Toc165152538)

[11. REFERENCIAS 32](#_Toc165152539)

[Anexo 1 – Actividades y sus resultados 36](#_Toc165152540)

# RESUMEN

En la presente investigación se recopilará, almacenará y procesarán variables climatológicas y de calidad del aire, para posteriormente visualizar dentro de un grupo de puntos seleccionados en un mapa, cuál sería el mejor lugar para ubicar una granja solar. Para lo cual, se propondrá un modelo LSTM que evaluará la eficiencia de los panes solares, bajo el efecto de las variables climatológicas, calidad del aire e irradiancia solar.

# PALABRAS CLAVE

Eficiencia de paneles solares, variables climatológicas, calidad del aire, irradiancia solar, machine learning.

# INTRODUCCIÓN

La energía solar se ha convertido en una fuente de energía renovable cada vez más importante en la búsqueda de un futuro sostenible como se menciona en (Hassan et al., 2024). Sin embargo, la eficiencia de los paneles solares puede verse afectada por diversos factores, como las variables climatológicas y la calidad del aire como se menciona en (Ahmed et al., 2023; Chanchangi et al., 2021; Saad et al., 2022; Vidyanandan, 2017a).

Por lo anterior, esta investigación tiene como objetivo principal determinar la influencia de las variables climatológicas y de la calidad del aire en la eficiencia de los paneles solares, con el fin de lograr una mejor ubicación de las granjas solares y aumentar la producción de energía renovable. Para alcanzar este objetivo, se utilizará un enfoque metodológico mixto que combina:

* Análisis cuantitativo: Recopilación de datos de variables climatológicas y de calidad del aire de diferentes ubicaciones mediante APIs de terceros.
* Desarrollo de un software en la nube: Creación de una herramienta de fácil uso que permita a los usuarios seleccionar puntos en un mapa y visualización de cual(es) de ellos tiene una eficiencia más cercana a valores teóricos.
* Análisis cualitativo: Revisión de la literatura científica sobre la influencia de las variables climatológicas y de la calidad del aire en la eficiencia de los paneles solares.

Esta investigación, se espera que contribuya a la lucha contra el cambio climático, la reducción de costos debido a una apropiada ubicación de las granjas solares para lograr una mejor eficiencia y aprovechamiento del recurso solar y el desarrollo de un futuro energético más sostenible.

Finalmente, es importante tener en cuenta que esta investigación tiene algunas limitaciones, como la disponibilidad y costos de datos insumo y posiblemente la generalización del modelo.

# PROBLEMA DE INVESTIGACIÓN

## PLANTEAMIENTO DEL PROBLEMA

En un mundo en transición energética, donde cada vez más surge la necesidad de instalar fuentes alternativas de energía que sustituyan nuestras fuentes convencionales fósiles para tratar de detener el cambio climático, vemos que se ha incrementado la instalación de generación solar, eólica e hídrica, siendo la primera la más popular debido a su facilidad de instalación y capacidad de generación distribuida y aislada.

No obstante, variables exógenas como la cantidad de sol recibida, la temperatura, la presión y la humedad según (Vidyanandan, 2017a), y la suciedad según (Chanchangi et al., 2021) y (Al Jassasi et al., 2023) pueden afectar significativamente la generación de potencia activa utilizable y por lo tanto afectar directamente la eficiencia de los paneles solares. Es por ello que dichas variables exógenas, toman relevancia en la correcta ubicación de los paneles solares como se menciona en (Chintapalli et al., 2020). y por lo tanto, cada vez más, encontramos en la literatura estudios de los efectos climatológicos en zonas específicas como se menciona en (Ahmed et al., 2023; Al Jassasi et al., 2023; Chanchangi et al., 2021)

De no contar con estudios sobre la incidencia de las variables climatológicas y la contaminación ambiental, que soporten la instalación de generación solar en lugar determinado, se podría llegar a escenarios donde su instalación aparentemente es viable, pero que con el pasar del tiempo y en un muy corto plazo, la eficiencia del panel se puede degradar rápidamente haciendo inviable la generación y por lo tanto generar no solo pérdidas económicas, sino también la desconexión de esas zonas que se pretendían conectar mediante generación distribuida.

Dependiendo de los datos climatológicos disponibles y provenientes de APIs reconocidas a nivel mundial ([Solcat](https://solcast.com/data-specifications), [OpenWeather](https://openweathermap.org/), [WeaterApi](https://www.weatherapi.com/)…), se pretende construir un modelo que estudie el comportamiento de la eficiencia de los paneles solares, en una ubicación preseleccionada en Colombia, bajo el efecto de las variables climatológicas seleccionadas así como la contaminación ambiental y soportando su computación en la nube, con lo cual se facilitará el estudio de la ubicación geográfica bajo estudio.

## FORMULACIÓN DEL PROBLEMA

¿Cómo evaluar la eficiencia de los paneles solares, basado en variables climatológicas, calidad del aire e irradiancia?

## SISTEMATIZACIÓN DEL PROBLEMA

A partir de la formulación del problema, se plantean las siguientes preguntas para el desarrollo de este trabajo de investigación:

¿Cómo debería ser la descripción funcional del modelo de evaluación y visualización, basado en variables climatológicas, calidad del aire e irradiancia?

¿Cómo debería ser la descripción estructural del modelo de evaluación y visualización basado en variables climatológicas, calidad del aire e irradiancia?

¿Cómo se puede validar, que el modelo propuesto satisface la predicción de la eficiencia, a partir de los datos observados y los datos predichos?

# OBJETIVOS

## OBJETIVO GENERAL

Diseñar un modelo de evaluación de la eficiencia de los paneles solares, basado en variables climatológicas, calidad del aire e irradiancia.

## OBJETIVOS ESPECÍFICOS

* Definir la descripción funcional del modelo de evaluación de la eficiencia de los paneles solares, basado en variables climatológicas, de calidad del aire e irradiancia.
* Definir la estructural del modelo de evaluación de la eficiencia de los paneles solares, basado en variables climatológicas, de calidad del aire e irradiancia.
* Validar el modelo propuesto, mediante la comparación de datos observados versus el resultado de la predicción.

# JUSTIFICACIÓN

## JUSTIFICACIÓN TEÓRICA

Esta investigación busca comprender la relación causal entre las variables climatológicas y de la calidad del aire, con la eficiencia de los paneles solares. Ya que diversos estudios previos, han demostrado que el rendimiento de los paneles solares se ve afectado, no solo por las condiciones climáticas como se menciona en (Saad et al., 2022) y (Vidyanandan, 2017a), sino como también por la calidad del aire como se menciona en (Ghazi & Ip, 2014) y (Mejia et al., 2014).

Es por ello que el objetivo principal, es ampliar el conocimiento actual sobre este tema, al determinar cómo las variables climatológicas y de la calidad del aire afectan la eficiencia de los paneles solares.

Por su parte, los resultados de la investigación pueden aportar información valiosa para la optimización de la ubicación de granjas solares y para comprender por qué la instalación en el pasado de alguna capacidad de generación no cumplió con lo que se esperaba.

## JUSTIFICACIÓN METODOLÓGICA

Se utilizará un diseño de investigación cuantitativo, para lo cual se recopilarán los datos de las variables climatológicas y de calidad del aire de diferentes ubicaciones, mediante el uso de APIs de terceros como SolCast y OpenWeather. La data, será almacenada para analizar la correlación entre cada una de las variables y poder determinar la conveniencia o no de las mismas, como variables de ingreso al modelo, pero no sin antes, realizar un análisis de cómo está distribuida la muestra, usando estadística descriptiva

Una vez se cuente con datos limpios, podría ser necesario aplicar un preprocesamiento de los mismos como se menciona en (Malvoni et al., 2017), para lograr una mejor predicción en términos de la eficiencia y cuyo resultado serán las variables que, mediante un análisis estadístico, nos permitirá llegar a un modelo que logre capturar las relaciones entre las variables climatológicas, de calidad del aire e irradiancia con respecto a la eficiencia de los paneles solares. No obstante, esta investigación tiene 2 grandes limitantes, la disponibilidad y costo de los datos, y que el modelo propuesto puede no ser general para todas las ubicaciones.

Finalmente, se generará un software en la nube de muy fácil uso, cuya interacción con el usuario sea un mapa, en el cual se puedan seleccionar una cantidad finita de puntos, para los cuales se aplicará todo el proceso descrito anteriormente y que, como resultado, muestre dentro de los puntos seleccionados, cuál es el más viable para instalar una cierta capacidad energética solar, basado en el histórico y predicción de las variables de calidad del aire y climatológicas. Esperando así que este software pueda ser el insumo de futuras investigaciones en el campo.

## JUSTIFICACIÓN PRÁCTICA

Esta investigación tiene el potencial de contribuir en tres áreas principales:

* Optimización de la ubicación de granjas solares: Al comprender la relación entre las variables climatológicas, la calidad del aire y la eficiencia de los paneles solares, la investigación puede ayudar a determinar los mejores lugares para instalar granjas solares. Esto permitirá maximizar la producción de energía, reducir los costos de generación y minimizar los impactos ambientales.
* Compresión de una baja eficiencia de instalaciones existentes: La investigación busca comprender por qué algunas granjas solares no funcionan tan bien como se esperaba, lo cual nos conduciría a mejorar el diseño y la operación, así como identificar si alguna de estas variables está afectando la generación esperada.
* Finalmente, beneficios económicos y sociales: La investigación puede generar nuevos empleos, puede hacer la industria solar más eficiente y sostenible, así como contribuir a la lucha del cambio climático y mejorar la calidad del aire, en las zonas donde se instalen las granjas.

# MARCO DE REFERENCIA

## MARCO CONCEPTUAL

* + 1. PANELES SOLARES

Para entender que es un panel solar, es necesario adentrarnos en su unidad mas pequeña de construcción; las celdas fotovoltáicas (Fig 1. Parte derecha superior), las cuales según (Vidyanandan, 2017b) son dispositivos semiconductores que convierten la energía de la luz solar captada por unidad de área, en electricidad. Y usualmente, un panel solar esta compuesto de 48, 60 o 72 de estas celdas, de manera que al ser connectadas secuencial o paralelamente, aumentan el área de recoleccion de irradiancia y por lo tanto, se logra aumentar la potencia de salida.

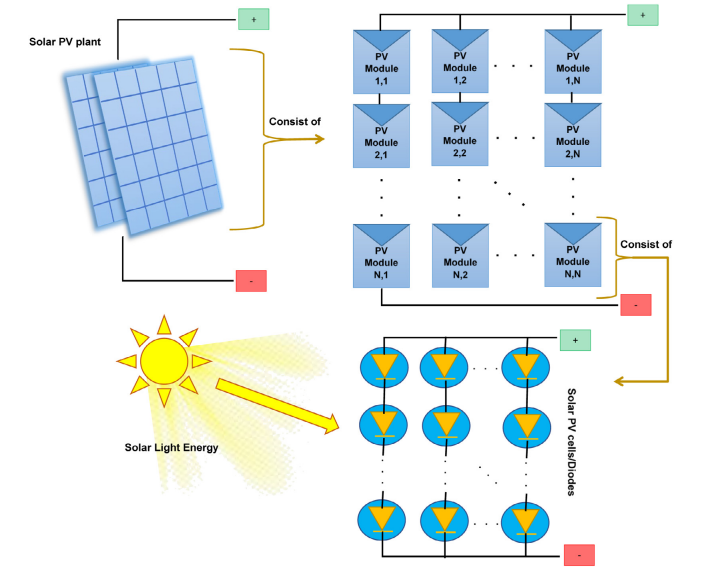


Fig 1. Panales fotovoltaicos (Abdel-Basset et al., 2022)

#### Celda Fotovoltaica

La composición interna de una celda fotovoltaica o celda solar se muestra en la Fig 2.a y su funcionamiento acorde a (Vidyanandan, 2017b) es el siguiente; cuando los fotones golpean la unión p-n Fig 2.b, los electrones se liberan de la banda de valencia y se elevan a la banda de conducción superando la energía de banda prohibida, Eg. La banda de valencia es el rango más alto de energía de los electrones, en el que estos no son libres de moverse y la banda de conducción es el rango más bajo de energía de electrones en el que pueden moverse libremente.

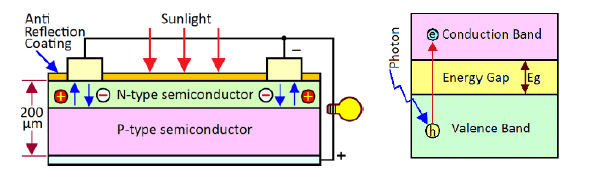


Fig 2. (a) izquierda vista interior de una celda, (b) derecha bandas de energía (Vidyanandan, 2017b)

En el caso del silicio, la banda prohibida es de 1.12 eV. Eso significa que los fotones con un nivel de energía de 1.12 eV o superior generarán corriente en la celda, mientras que los fotones con un nivel de energía por debajo de 1.12 eV pasarán sin ser absorbidos. La elevación del electrón de la banda de valencia a la banda de conducción crea un hueco en la banda de valencia. Los electrones libres en la banda de conducción y los huecos en la banda de valencia son los responsables del flujo de corriente cuando la celda FV se conecta a una carga.

#### Tipos de celdas

Según (Hegedus & Luque, 2003) existen 4 categorías principales (Fig 3) en las cuales se agrupa la evolución de las celdas fotovoltaicas basado en sus materiales de construccion, donde estas categorías son mayormente conocidas como generaciones de la tecnología solar.

Primera Generación: Acorde a (Pastuszak & Węgierek, 2022) son celdas basadas en silicio ya sea mono-, poli- o multi-cristalino, así como junturas III-V(GaAs) y fueron las primeras en ser introducidas al mercado, debido a su relativa alta eficiencia y durabilidad, con una penetración poco superior al 80%.

Segunda Generación: Según (Pastuszak & Węgierek, 2022), esta segunda generación fue diseñada para ser un reemplazo de bajo costo a las celdas cristalinas de silicio, las cuales están basadas en CdTe, seleniuro de galio y cobre (CIGS) o silicio amorfo. Estas son ideales para aplicaciones flexibles sin embargo su eficiencia se reduce.

Tercera Generación: La tercera generación de células solares, la cual incluyen las células tándem, de perovskita, sensibilizadas por colorante, orgánicas y conceptos emergentes, representa una amplia gama de enfoques. Abarca, según (Pastuszak & Węgierek, 2022) desde sistemas económicos de baja eficiencia (celdas solares orgánicas, sensibilizadas por colorante) hasta sistemas costosos de alta eficiencia (celdas multiunión III-V) para aplicaciones que van desde la integración en edificios hasta aplicaciones espaciales. Las células solares de tercera generación a veces se denominan 'conceptos emergentes' debido a su baja penetración en el mercado.

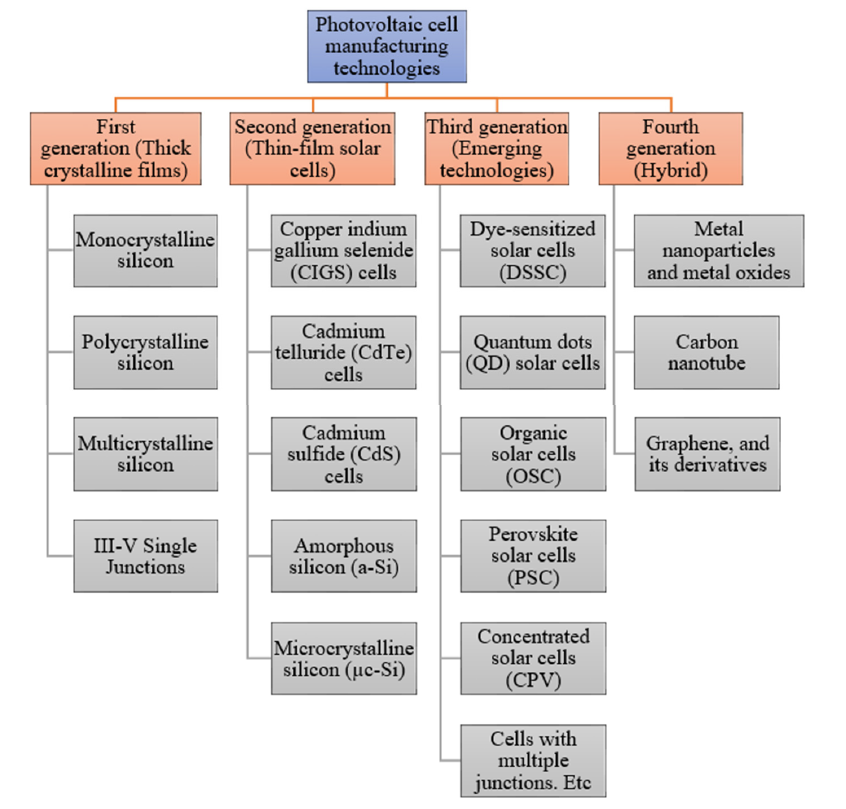


Fig 3. Tipos de celdas solares según (Pastuszak & Węgierek, 2022)

Cuarta Generación: De acuerdo con (Pastuszak & Węgierek, 2022), las celdas fotovoltaicas de cuarta generación, también se conocen como células inorgánicas híbridas porque combinan el bajo costo y la flexibilidad de las películas delgadas de polímero, con la estabilidad de las nanoestructuras orgánicas como nanopartículas metálicas, óxidos metálicos, nanotubos de carbono, grafeno y sus derivados.

Cada una de las generaciones de la tecnología solar, trae consigo una eficiencia promedio que es inherente a los materiales de fabricación la cual se condensa en la Fig 4. Es importante resaltar que dicha eficiencia varía actualmente del 13% hasta casi el 48%.

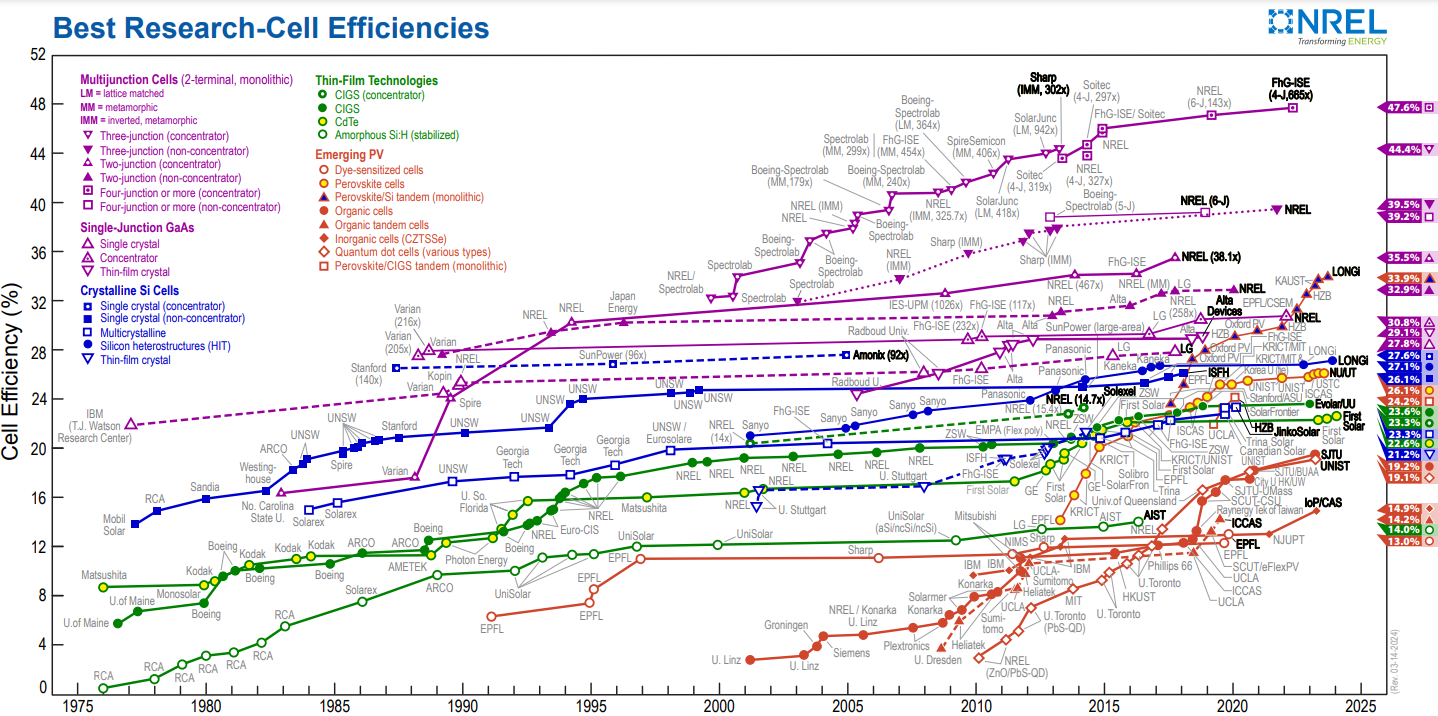


Fig 4. Eficiencia típica de las celdas fotovoltaicas (NREL Transforming ENERGY, n.d.)

#### Modelo Eléctrico

En la industria solar, se tiene diferentes modelos eléctricos para representar el comportamiento de los paneles solares. Que como es explicado detalladamente en (Lameirinhas et al., 2022) se tiene el modelo 1M3P, lo que significa 1 modelo con 3 parámetros, siendo esta la versión más simple, la cual no representa la relación que los paneles poseen con la irradiación. Es por ello que su siguiente versión 1M5P que es entendido como 1 modelo de 5 parámetros (Fig 5), es la que usaremos para llegar a la definición de la eficiencia.

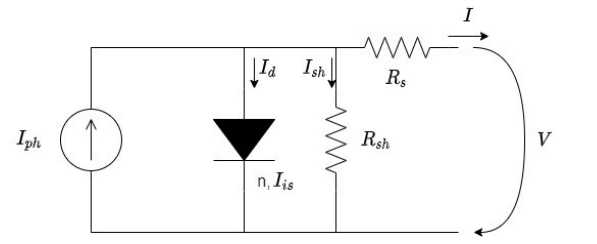


Fig 5. Circuito equivalente 1M5P (Lameirinhas et al., 2022)

Como se expresa en (Lameirinhas et al., 2022), la curva mostrada en la Fig 6 muestra la relación entre la corriente de corto circuito y el voltaje de circuito abierto en azul del modelo eléctrico de la Fig 5, y se define el punto de operación de la celda solar Q, como el punto en el cual se logra máxima transferencia de potencia. Dicho punto es muy importante, ya que los fabricantes de las celdas, entregan estas curvas de comportamiento basados en condiciones de prueba estándar (STC – Standard Test Conditions) realizadas con una irradiancia de , temperatura de y masa de aire de 1.5 de espectro global, pero adicionalmente estas curvas cambian a medida que la resistencia de derivación (Fig 5, Señalada como , shunt resistance) lo hace.

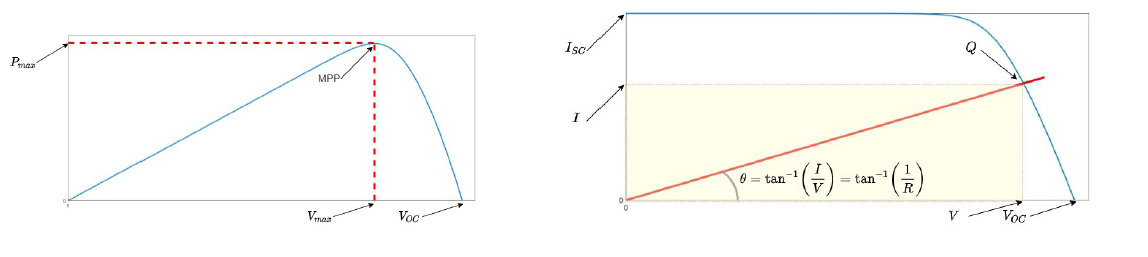


Fig 6. Ejemplo de caracterización de una celda solar (en azul) (izquierda potencia vs voltaje, derecha corriente de corto circuito vs voltaje de circuito abierto) y la definición del punto de máxima potencia Q (rojo) (Lameirinhas et al., 2022)

Finalmente, la eficiencia de una celda solar según (Lameirinhas et al., 2022), es definida mediante la siguiente ecuación, en la cual , representa la potencia máxima calculada a partir de la corriente máxima ( y voltaje máximo del punto Q, y la potencia incidente (, la cual es calculada a partir de la y el área activa del panel .

* + 1. IRRADIANCIA

Acorde a (Vidyanandan, 2017) la radiación solar representa todo el espectro electromagnético, el cual está compuesto por rayos gamma de alta energía, seguidos por rayos X, ultravioleta, luz visible, infrarrojo, microondas y ondas de radio de baja energía (Fig 7). Estas radiaciones se describen como un flujo de partículas sin masa llamadas fotones, que viajan a la velocidad de la luz y contienen una cantidad definida de energía. La diferenciación entre las distintas radiaciones se basa en la cantidad de energía que poseen los fotones.

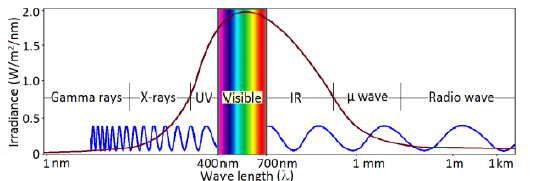


Fig 7. Espectro electromagnético

La energía de los fotones, medida en electronvoltios (eV), está relacionada con la longitud de onda de la radiación. Los rayos gamma tienen los fotones más energéticos, mientras que las ondas de radio tienen fotones con energías bajas. La longitud de onda de la luz visible varía entre 380 y 720 nm. La luz con una longitud de onda más corta (por ejemplo, la luz azul) contiene fotones de alta energía, y la luz con una longitud de onda más larga (por ejemplo, la luz roja) contiene fotones de baja energía.

#### GHI

Acorde (Prado-Rujas et al., 2021), la irradiancia global horizontal (GHI – Global Horizontal Irradiance) es la cantidad total de radiación de onda corta recibida por una superficie horizontal. Por lo tanto, su unidad de medición es y se puede expresar en términos de la irradiancia horizontal difusa (DHI – Diffused Horizontal Irraciance) y la irradiancia normal directa (DNI – Diffused Normal Irradiance) mediante la siguiente ecuación, donde z representa el ángulo cenit del sol.

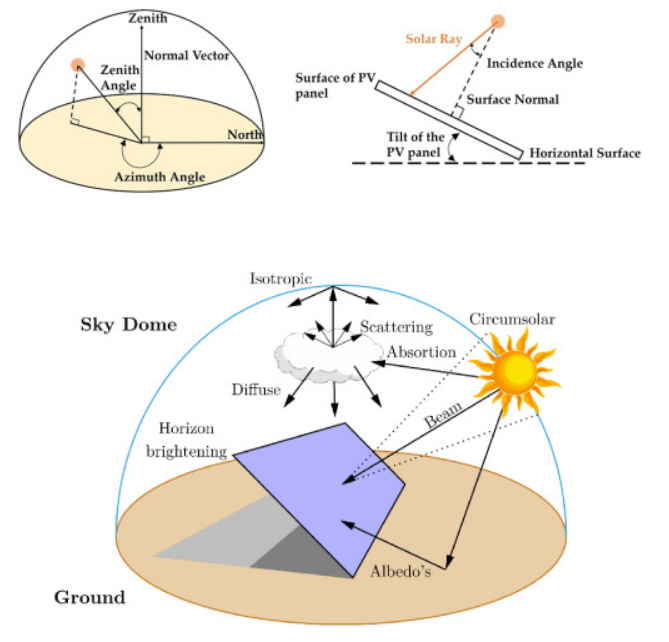


Fig 8. Componentes de la irradiancia solar en una superficie inclinada (Kosmopoulos, 2024)

* + 1. CALIDAD DEL AIRE (Sagheer et al., 2024)

La polución es definida como la presencia de sustancias en el ambiente que tienen efectos adversos en la salud humana y en los ecosistemas. De la cual se deriva la contaminación del aire al ser una mezcla partículas y gases en el ambiente.

Los contaminantes del aire se pueden clasificar en contaminantes primarios, los cuales son emitidos directamente a la atmosfera dentro de los cuales se incluyen el material particulado, el monóxido de carbono, el dióxido de carbono y el dióxido de azufre. Pero también pueden ser contaminantes secundarios, los cuales se forman con interacciones con otros componentes y dentro de los cuales se tiene el ozono y derivados de los materiales particulados.

#### PM - Material Particulado

El PM es una mezcla compleja de partículas sólidas y materia gaseosa. Su composición varía según el tiempo, la ubicación, las condiciones ambientales y las fuentes de emisión y se clasifican por diámetro aerodinámico en: partículas gruesas con un diámetro de 2,6 a 10 micrómetros (µm), partículas finas con un diámetro de 0,1 a 2,5 µm y partículas ultrafinas (UFP o ) con un diámetro <0,1 µm. PM2,5 y PM10 son las fracciones de MP más estudiadas en relación con los efectos adversos para la salud, siendo PM2,5 la más relevante para las enfermedades cardiovasculares.

* + 1. ALGORITMOS DE PREDICCIÓN

#### ANN - Redes neuronales artificiales

Acorde (Chandel et al., 2023) la estructura de una RNA red neuronal artificial (ANN – Artificial Neuronal Network) consisten de 3 capas; entrada, oculta y salida, con un numero determinado de neuronas como se muestra en Fig 9.

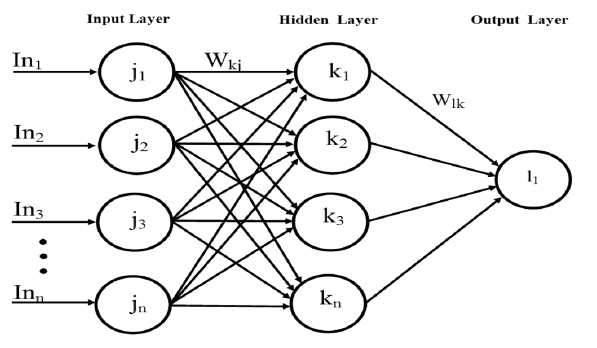


Fig 9. Estructura de una RNA según (Chandel et al., 2023)

Y según (Chandel et al., 2023) estas capas están conectada entre sí, usando pesos que mapean la entrada a una salida deseada, usando la ecuación siguiente, donde es el número de entradas, es el peso de la matriz, es la entrada de la capa anterior, B es un bias and es la activación de la función.

De la forma descrita anteriormente, una RNA imita el proceso de un cerebro humano. Para lo cual en modo de entrenamiento las muestras alimentan las entradas y al principio del entrenamiento, los pesos son aleatoriamente inicializados para lograr igualmente una salida arbitraria. Luego, el valor obtenido en la salida es comparado al valor deseado, usando funciones de costo como RSME, MSE, MAE entre otras para ser propagado hacia atrás en función de cambiar nuevamente los pesos para lograr acercar el valor de la salida con el valor real.

#### GRU - Redes neuronales recurrentes con puerta

Una GRU (Gate Recurrent Unit) como es mencionado en (Cho et al., 2014) es un tipo especial de RNA que nos sirve para procesar información secuencial y consiste en 2 puertas (Fig 10) una de reset , que decide si se incluye la información del estado anterior y otra de actualización , que decide qué cantidad de información del estado anterior es tenida en cuenta en el estado actual. Sí el valor de la puerta de reset es 0, entonces el estado oculto actual ignora la información del estado anterior.

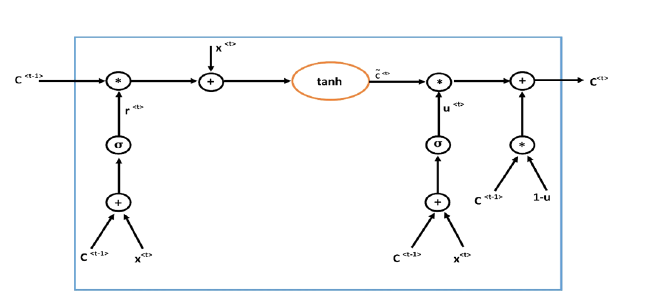


Fig 10. Arquitectura básica de GRU

#### LSTM - Redes neuronales recurrentes de largo plazo corto

Una LSTM (Long short-term memory base network) es un tipo especial de RNAs propuesto inicialmente por (Hochreiter & Schmidhuber, 1997) para lograr aprender dependencias de largo plazo entre los estados. LSTM consiste en varias puertas; puerta de ingreso , puerta de salida, puerta de olvido y estado de la celda, como se muestra en la Fig 11.

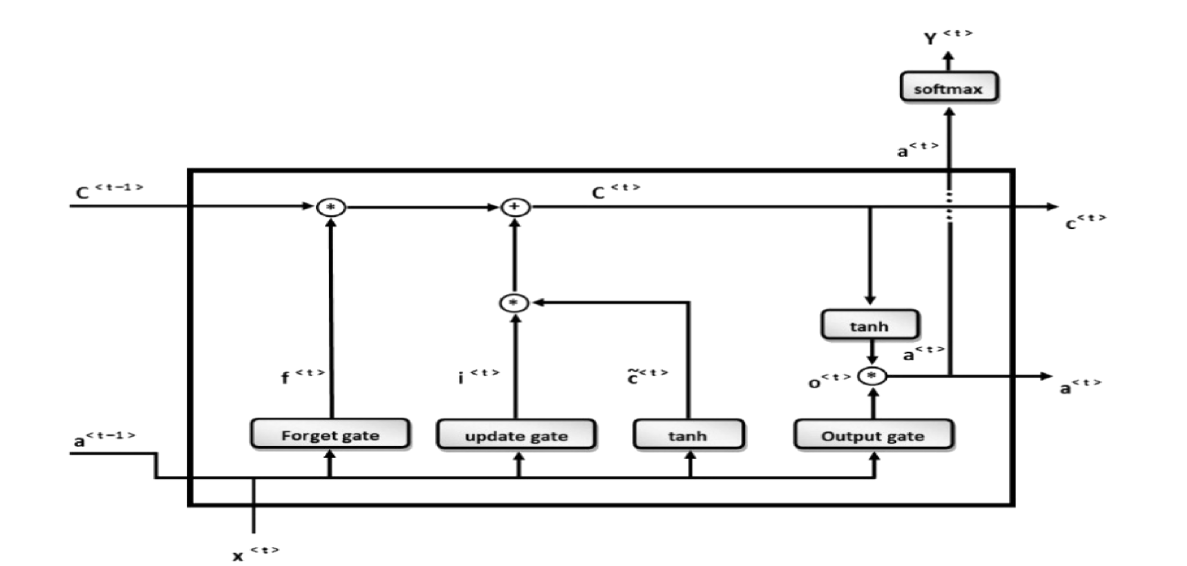


Fig 11. Unidad de LSTM (Chandel et al., 2023)

#### CNN – Redes Neuronales Convolusionales

Una CNN (Convolutional Neural Network) red neuronal convolucional, de acuerdo a (Chandel et al., 2023) procesa data en forma de arrays como imágenes o textos y su arquitectura interna se muestra en Fig 12. En la cual se pueden ver una secuencia de capas convolucionales y de agrupamiento, seguidas de capas completamente conectadas en las últimas etapas. La salida de cada capa convolucional o de agrupamiento se pasa a una función de activación ReLU.

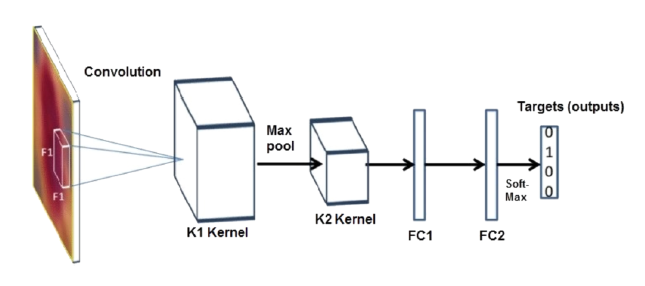


Fig 12. Arquitectura básica de una CNN

## MARCO TEÓRICO

De acuerdo con (Li et al., 2022), a medida que se generaliza el uso de la energía fósil tradicional, los problemas del agotamiento de estas fuentes y la contaminación ambiental se vuelven cada vez más graves. Por eso, la energía fotovoltaica (FV) ha ganado popularidad a nivel mundial gracias a sus ventajas: es limpia, no contaminante y favorece una fácil distribución. Es por ello y como menciona (Chandel et al., 2023), que las centrales fotovoltaicas se han vuelto más competitivas en comparación con las plantas de combustibles fósiles en los últimos años. Sin embargo, la naturaleza fluctuante de la irradiancia solar y las condiciones climáticas y geográficas hacen que la generación de energía sea impredecible, afectando así el rendimiento de las plantas solares y la red eléctrica.

Por su parte, la predicción de la energía fotovoltaica es una herramienta vital para los operadores y gestores de plantas solares, ya que les ayuda a evitar penalizaciones por las diferencias entre la energía fotovoltaica real y la deseada, así como evaluar en el momento de planeación, si un lugar u otro es apropiado para la instalación de las plantas solares. En este sentido, se encuentra gran cantidad de modelos de predicción, basado en datos climáticos (temperatura, velocidad del viento, humedad relativa, y presión del aire), irradiancia y suciedad de los paneles, para diferentes zonas del planeta, como por ejemplo (Malvoni et al., 2017) en el cual se realiza un modelo de predicción de generación de potencia, usando descomposición wavelet para Salento – Italia, o como en (Sahin et al., 2023), en el cual se propone un modelo que usa una FFNN (Feedforward neural network) red neuronal prealimentada, para predecir la eficiencia de una planta solar basado en variables climáticas e irradiancia en Igdir – Turkía. También, como se menciona en (Chanchangi et al., 2021), donde lograron demostrar que la eficiencia de los diferentes tipos de panales solares en el norte de Nigeria, se vería fuertemente impactada por la cantidad de polvo en su superficie, pero que a su vez la velocidad del viento y las lluvias tuvieron un efecto de limpieza de la suciedad y que por lo tanto, se debían efectuar esquemas de limpieza de los paneles para prevenir la degradación del desempeño. Por otra parte, en el desierto del Sahara ubicado en Argelia, según se menciona en (Memiche et al., 2020) se analizó el rendimiento y la contaminación causada por el polvo en el rendimiento de los paneles, encontrando que existe una relación lineal entre la perdida de potencia y la acumulación de suciedad.

Es por ello, que el autor quisiera ampliar el conocimiento existente en ese sentido, proponiendo un modelo de predicción que use un algoritmo LSTM, que incluya además de las variables climatológicas e irradiancia, las variables asociadas a la contaminación ambiental como lo son el material particulado P2.5, P10 y el monóxido de carbono principalmente. Sn embargo, las APIs que se usaran en este trabajo también poseen información de otros contaminantes como lo son: el monóxido y dióxido de nitrógeno, amonio y dióxido de azufre, así como la cantidad de ozono en la atmósfera, las cuales serán tenidas en cuenta durante el estudio de correlación para determinar si deberían ser incluidas en el estudio.

## MARCO ESPACIAL

El autor, motivado por el gran interés en predecir la generación de energía mediante paneles solares en una ubicación específica de Colombia Fig 13, y considerando que la calidad de los datos disponibles sobre este tema no es la ideal, es necesario buscar un lugar en el mundo que comparta características climatológicas y de contaminación ambiental similares a las de Colombia, y que a su vez, cuente con datos de generación de energía solar de alta calidad. Esto permitirá realizar una evaluación más precisa y confiable del potencial de la energía solar en el contexto colombiano.

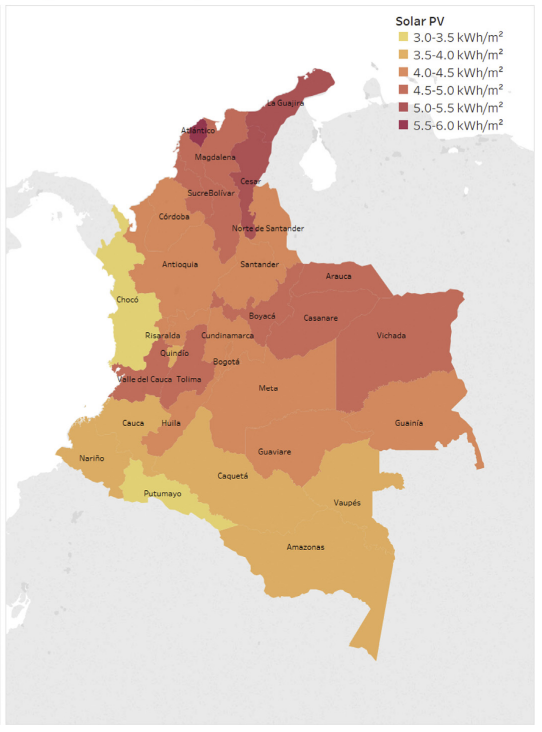


Fig 13. Irradiancia promedio por departamentos en Colombia (López et al., 2020)

## MARCO TEMPORAL

Si bien no existe un consenso definitivo sobre la cantidad de años de datos necesarios para entrenar un modelo similar al propuesto, se consideran factores como el costo de adquisición de datos a través de APIs y la evidencia de estudios previos como se menciona en (Jeong et al., 2023), (Liang et al., 2023) y (Gonçalves et al., 2023) en los cuales utilizan 1, 2 y hasta 4 años de datos respetivamente en algoritmos de predicción de corto y mediano plazo. En base a esto, se estima inicialmente que 2 años de datos para los puntos seleccionados son suficientes para obtener predicciones significativas de la eficiencia. Sin embargo, esta estimación inicial se evaluará y ajustará en función de los resultados obtenidos durante la investigación, incluyendo un análisis de sensibilidad para determinar el impacto de diferentes cantidades de datos en el rendimiento del modelo.

# HIPÓTESIS

1. Al incluir datos de calidad del aire, adicionales a datos climatológicos e irradiancia en un modelo LSTM, se logrará predecir la eficiencia de los paneles solares.
2. Al incluir datos de calidad del aire, adicionales a datos climatológicos e irradiancia en un modelo LSTM, se logrará la predicción de la eficiencia, tal que se reduzca la brecha entre el entrenamiento y la realidad, para que esta solo sea de un 15% máximo cuando se tiene un horizonte temporal de hasta 2 meses (Mid-term horizon).

**Hipótesis de arranque por construcción** (basada en teorías previas) **y analogía sustantiva** (comparación con otro modelo que no incluía variables de calidad del aire), **ostensiva no observacional** (variables no pueden ser observables) **teorética** (basada en teoría científica) **y profunda fenomenológica** (relación in-out, sin indicar el porqué)

Estas 2 hipótesis, se basan en el hecho de que hay variables relevantes al proceso de predicción, que no fueron tenidos en cuenta, así como la calidad de los datos.

# METODOLOGÍA

El proyecto de investigación se realizará en 3 fases, cada una de las cuales aportará significativamente en la consecución de los objetivos específicos (Ver anexo 1 – Actividades y sus resultados).

**Fase I. Revisión de la literatura, definición del alcance y de la fuente de datos.**

Objetivo: En esta fase se tiene un acercamiento a los datos y por lo tanto, se determinará cuál será su fuente, comportamiento y correlación para validar si su dimensionalidad es la correcta, para posteriormente evaluar el desempeño y seleccionar el algoritmo apropiado para la predicción que según la hipótesis puede ser LSTM. Por lo anterior, se utilizará una metodología estadística, aplicando estadística descriptiva e inferencial.

* Leer y analizar distintos artículos científicos sobre predicción de eficiencia de paneles solares y algoritmos usados para tal fin.
* Buscar diferentes instituciones, compañías o sitios web que ofrezcan datos correspondientes a variables climatológicas, calidad el aire e irradiancia a nivel mundial y que dicha información sea expuesta a través de API rest.
* Recopilar los datos de las variables climatológicas, calidad el aire, irradiancia y eficiencia en una base de datos.
* Preprocesamiento de los datos (Eliminación de datos atípicos, llenado de datos faltantes, Normalización)
* Realizar estadística descriptiva sobre los mismos.
* Realizar un análisis de correlación sobre los datos, para reducir la dimensionalidad si aplica.
* Evaluar y seleccionar diferentes algoritmos de predicción.

**Fase II. Desarrollo del modelo.**

Objetivo: Esta parte se centra en actividades propias de la construcción de software, por lo cual se usará metodología ágil. Y, debido a que el investigador trabajará solo, la más apropiada es Kanban.

* Realizar diagrama de arquitectura de la solución propuesta, incluyendo vista de casos de uso, componentes y secuencia.
* Configuración de repositorios en GitLab
* Realizar una cotización de los componentes requeridos de la nube
* Configuración de los componentes de nube requeridos.
* Seleccionar y configurar servicio SaaS para manejo de mapas.
* Desarrollo de microservicio BE integrador, para extraer la data del sitio seleccionado y llevarla a nuestro repositorio.
* Generación de script en terraform (CI/CD)
* Desarrollo de microservicio FE, integrado con servicio SaaS para manejo de mapas.
* Realizar pruebas de integración del sistema propuesto.

**Fase III. Validación, aplicación del modelo a un caso real y documentación.**

Objetivo: En la fase final, se validará el modelo desarrollado al aplicarlo a un caso de estudio real, para posteriormente: documentar exhaustivamente el proceso, los resultados y difundir los hallazgos a la comunidad científica. Es por ello que se usará una metodología cuantitativa ya que la idea es poder encontrar, para luego exponer el conocimiento encontrado.

* Recopilar datos observados de la eficiencia de paneles solares para el caso de estudio propuesto.
* Realizar las conclusiones a las que haya lugar al respecto de la investigación
* Realizar pruebas para el caso de estudio propuesto.
* Realizar paper con resultados finales de la investigación.
* Realizar documento definitivo de la investigación.

Finalmente, en la Tabla 1. Avance ponderado de objetivos a través de las fases se muestra el detalle y porcentaje de avance respecto a los objetivos específicos con relación a la fase en la que se encuentra en proyecto.

Tabla 1. Avance ponderado de objetivos a través de las fases

|  |  |  |  |
| --- | --- | --- | --- |
| **Objetivo específico** | **Fase I** | **Fase II** | **Fase III** |
| Definir la descripción funcional del modelo de evaluación de la eficiencia de los paneles solares, basado en variables climatológicas, de calidad del aire e irradiancia. | 60 | 25 | 15 |
| Definir la estructural del modelo de evaluación de la eficiencia de los paneles solares, basado en variables climatológicas, de calidad del aire e irradiancia. | 25 | 60 | 15 |
| Validar el modelo propuesto, mediante la comparación de datos resultado de la predicción y los datos observados. | 10 | 30 | 60 |

# IMPACTO Y RESULTADOS ESPERADOS

Sobre la Tabla 2 se muestra el impacto de los resultados que tendrá este trabajo de investigación, indicando su beneficiario.

Tabla 2. Impacto de los resultados

|  |  |  |
| --- | --- | --- |
| **Resultado/Producto esperado** | **Indicador** | **Beneficiario** |
| Optar por un título como Maestría en Ciencias de la Información y las Comunicaciones con énfasis en software | Documento de Trabajo de grado, acta sustentación | Luis Carrillo Mejía |
| Aceptación por parte del director | Que el proyecto sea ingresado como experiencia dentro del currículo | Elvis Eduardo Gaona García |
| Aceptación por parte del codirector | Que el proyecto sea ingresado como experiencia dentro del currículo | El codirector del proyecto |
| Compartir el trabajo de investigación | Publicación de los resultados finales en alguna revista de ingeniería. | La comunidad científica con área de interés fuentes fotovoltaicas de generación de energía |

# RECURSOS Y PRESUPUESTO

El presupuesto debe cubrir en detalle los rubros de recursos humanos, infraestructura física, equipos, materiales, viajes, bibliografía y otros. Cuando sea del caso se debe dimensionar la dedicación o duración en el tiempo del recurso y su costo. Se debe especificar la fuente de los recursos (personal, universidad, institución, etc.), en todos los casos se debe valorar cada ítem.

# CRONOGRAMA

En esta sección se relacionan el orden de ejecución de las actividades a realizar y los resultados a obtener en función de tiempo. Debe detallarse con bastante claridad el itinerario a seguir en el cumplimiento de los objetivos previstos en la investigación. En el cronograma, que es necesario ajustar dado los periodos de tiempo señalados por el programa, se explicitan aquellas actividades y tareas que acometerá el investigador. El cronograma es un indicador de coherencia y articulación de la investigación y debe estar relacionado directamente con las actividades descritas en la metodología.

# LIMITACION

Se debe especificar si el proyecto tiene alguna limitación y en qué forma se disminuiría o influiría en el alcance del proyecto y/o en el desarrollo de las actividades.

# REFERENCIAS

Abdel-Basset, M., Mohamed, R., Sharawi, M., Abdel-Fatah, L., Abouhawwash, M., & Sallam, K. (2022). A comparative study of optimization algorithms for parameter estimation of PV solar cells and modules: Analysis and case studies. *Energy Reports*, *8*, 13047–13065. https://doi.org/https://doi.org/10.1016/j.egyr.2022.09.193

Ahmed, M. S., Karal, R., Das, B. K., & Das, A. (2023). Experimental investigation of cooling, wind velocity, and dust deposition effects on solar PV performance in a tropical climate in Bangladesh. *Case Studies in Thermal Engineering*, *50*, 103409. https://doi.org/10.1016/J.CSITE.2023.103409

Al Jassasi, I. S., Al Hashmi, H. S., Al Humairi, A., Bulale, Y., Husain, A., Al-Azzawi, M., & Jung, P. (2023). Experimental investigation of the soiling effect on the PV systems performance and the cleaning intervals in Oman. *Solar Energy Advances*, *3*. https://doi.org/10.1016/j.seja.2023.100045

Chanchangi, Y. N., Ghosh, A., Baig, H., Sundaram, S., & Mallick, T. K. (2021). Soiling on PV performance influenced by weather parameters in Northern Nigeria. *Renewable Energy*, *180*, 874–892. https://doi.org/10.1016/j.renene.2021.08.090

Chandel, S. S., Gupta, A., Chandel, R., & Tajjour, S. (2023). Review of deep learning techniques for power generation prediction of industrial solar photovoltaic plants. *Solar Compass*, *8*, 100061. https://doi.org/10.1016/J.SOLCOM.2023.100061

Chintapalli, N., Sharma, M. K., & Bhattacharya, J. (2020). Linking spectral, thermal and weather effects to predict location-specific deviation from the rated power of a PV panel. *Solar Energy*, *208*, 115–123. https://doi.org/https://doi.org/10.1016/j.solener.2020.07.080

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724–1734). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1179

Ghazi, S., & Ip, K. (2014). The effect of weather conditions on the efficiency of PV panels in the southeast of UK. *Renewable Energy*, *69*, 50–59. https://doi.org/https://doi.org/10.1016/j.renene.2014.03.018

Gonçalves, R., Ribeiro, V. M., & Pereira, F. L. (2023). Variable Split Convolutional Attention: A novel Deep Learning model applied to the household electric power consumption. *Energy*, *274*, 127321. https://doi.org/https://doi.org/10.1016/j.energy.2023.127321

Hassan, Q., Algburi, S., Sameen, A. Z., Al-Musawi, T. J., Al-Jiboory, A. K., Salman, H. M., Ali, B. M., & Jaszczur, M. (2024). A comprehensive review of international renewable energy growth. *Energy and Built Environment*. https://doi.org/10.1016/J.ENBENV.2023.12.002

Hegedus, S. S., & Luque, A. (2003). Status, Trends, Challenges and the Bright Future of Solar Electricity from Photovoltaics. In *Handbook of Photovoltaic Science and Engineering* (pp. 1–43). John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1002/0470014008.ch1

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-term Memory. *Neural Computation*, *9*, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jeong, S.-G., Do, Q. V., & Hwang, W.-J. (2023). Short-term photovoltaic power forecasting based on hybrid quantum gated recurrent unit. *ICT Express*. https://doi.org/https://doi.org/10.1016/j.icte.2023.12.005

Kosmopoulos, P. (2024). Solar irradiance and exploitation of the Sun’s power. *Planning and Management of Solar Power from Space*, 1–20. https://doi.org/10.1016/B978-0-12-823390-0.00006-5

Lameirinhas, R. A. M., Torres, J. P. N., & de Melo Cunha, J. P. (2022). A Photovoltaic Technology Review: History, Fundamentals and Applications. *Energies*, *15*(5), 1–44. https://ideas.repec.org/a/gam/jeners/v15y2022i5p1823-d761979.html

Li, Z., Xu, R., Luo, X., Cao, X., Du, S., & Sun, H. (2022). Short-term photovoltaic power prediction based on modal reconstruction and hybrid deep learning model. *Energy Reports*, *8*, 9919–9932. https://doi.org/10.1016/J.EGYR.2022.07.176

Liang, L., Su, T., Gao, Y., Qin, F., & Pan, M. (2023). FCDT-IWBOA-LSSVR: An innovative hybrid machine learning approach for efficient prediction of short-to-mid-term photovoltaic generation. *Journal of Cleaner Production*, *385*, 135716. https://doi.org/https://doi.org/10.1016/j.jclepro.2022.135716

López, A. R., Krumm, A., Schattenhofer, L., Burandt, T., Montoya, F. C., Oberländer, N., & Oei, P. Y. (2020). Solar PV generation in Colombia - A qualitative and quantitative approach to analyze the potential of solar energy market. *Renewable Energy*, *148*, 1266–1279. https://doi.org/10.1016/J.RENENE.2019.10.066

Malvoni, M., De Giorgi, M. G., & Congedo, P. M. (2017). Forecasting of PV Power Generation using weather input data-preprocessing techniques. *Energy Procedia*, *126*, 651–658. https://doi.org/10.1016/j.egypro.2017.08.293

Mejia, F., Kleissl, J., & Bosch, J. L. (2014). The Effect of Dust on Solar Photovoltaic Systems. *Energy Procedia*, *49*, 2370–2376. https://doi.org/10.1016/J.EGYPRO.2014.03.251

Memiche, M., Bouzian, C., Benzahia, A., & Moussi, A. (2020). Effects of dust, soiling, aging, and weather conditions on photovoltaic system performances in a Saharan environment—Case study in Algeria. *Global Energy Interconnection*, *3*(1), 60–67. https://doi.org/10.1016/J.GLOEI.2020.03.004

NREL Transforming ENERGY. (n.d.). *NREL’s Best Research-Cell Efficiency Chart 2022*. Retrieved April 26, 2024, from https://www.nrel.gov/pv/cell-efficiency.html

Pastuszak, J., & Węgierek, P. (2022). Photovoltaic Cell Generations and Current Research Directions for Their Development. *Materials*, *15*. https://doi.org/10.3390/ma15165542

Prado-Rujas, I.-I., García-Dopico, A., Serrano, E., & Pérez, M. S. (2021). A Flexible and Robust Deep Learning-Based System for Solar Irradiance Forecasting. *IEEE Access*, *9*, 12348–12361. https://doi.org/10.1109/ACCESS.2021.3051839

Saad, B., El Hannani, A., Errattahi, R., & Aqqal, A. (2022). The impact of weather and panel characteristics features on photovoltaic power forecasting. *Materials Today: Proceedings*, *66*, 289–294. https://doi.org/https://doi.org/10.1016/j.matpr.2022.05.156

Sagheer, U., Al-Kindi, S., Abohashem, S., Phillips, C. T., Rana, J. S., Bhatnagar, A., Gulati, M., Rajagopalan, S., & Kalra, D. K. (2024). Environmental Pollution and Cardiovascular Disease: Part 1 of 2: Air Pollution. *JACC: Advances*, *3*(2), 100805. https://doi.org/10.1016/J.JACADV.2023.100805

Sahin, G., Isik, G., & van Sark, W. G. J. H. M. (2023). Predictive modeling of PV solar power plant efficiency considering weather conditions: A comparative analysis of artificial neural networks and multiple linear regression. *Energy Reports*, *10*, 2837–2849. https://doi.org/https://doi.org/10.1016/j.egyr.2023.09.097

Vidyanandan, K. V. (2017a). An Overview of Factors Affecting the Performance of Solar PV Systems. *Energy Scan (A House Journal of Corporate Planning, NTPC Ltd.)*, *27*, 2–8.

Vidyanandan, K. V. (2017b). An Overview of Factors Affecting the Performance of Solar PV Systems. *Energy Scan (A House Journal of Corporate Planning, NTPC Ltd.)*, *27*, 2–8.

# Anexo 1 – Actividades y sus resultados

Tabla 3. Actividades y sus resultados

|  |  |  |  |
| --- | --- | --- | --- |
| **Objetivo Específico** | **Metodología** | **Actividades** | **Resultado de la actividad** |
| Definir la descripción funcional del modelo de evaluación de la eficiencia de los paneles solares, basado en variables climatológicas, de calidad del aire e irradiancia. | En esta fase se tiene un acercamiento a los datos y, por lo tanto, se determina cuál será su fuente, comportamiento y correlación para validar si su dimensionalidad es la correcta, para posteriormente evaluar el desempeño y seleccionar el algoritmo apropiado para la predicción. Por lo anterior, se utilizará una metodología estadística aplicando estadística descriptiva e inferencial. | Leer y analizar distintos artículos científicos sobre predicción de eficiencia de paneles solares y algoritmos usados para tal fin. | Documento con resumen de al menos 20 referencias referentes del tema. |
| Buscar diferentes instituciones, compañías o sitios web que ofrezcan datos correspondientes a variables climatológicas, calidad el aire e irradiancia a nivel mundial y que dicha información sea expuesta a través de API rest. | Repositorio en Gitlab con las apis a ser usadas en el trabajo y un ejemplo de los payloads si los usa |
| Recopilar los datos de las variables climatológicas, calidad el aire, irradiancia y eficiencia en una base de datos. | Base de datos, con los datos relacionados a variables climatológicas, calidad del aire, irradiancia y potencia |
| Preprocesamiento de los datos (Eliminación de datos atípicos, llenado de datos faltantes, Normalización) | Base de datos con datos limpios |
| Realizar estadística descriptiva sobre los mismos. | Documento con análisis de estadística descriptiva |
| Realizar un análisis de correlación sobre los datos, para reducir la dimensionalidad si aplica. | Documento con variables relevantes de ingreso al modelo |
| Evaluar y seleccionar diferentes algoritmos de predicción. | Algoritmo para realizar la predicción definido |
| Definir la estructural del modelo de evaluación de la eficiencia de los paneles solares, basado en variables climatológicas, de calidad del aire e irradiancia. | Esta parte se centra en actividades propias de la construcción de software, por lo cual se usará metodología agil y debido a que el investigador trabaja solo, la más apropiada es Kanban. | Realizar diagrama de arquitectura de la solución propuesta, incluyendo vista de casos de uso, componentes y secuencia. | Diagramas UML de componente, casos de uso y secuencia |
| Configuración de repositorios en GitLab | Repositorio gitlab y callbacks configurados |
| Realizar una cotización de los componentes requeridos de la nube | Documento con calculadora de las diferentes nubes para los componentes requeridos |
| Configuración de los componentes de nube requeridos | Recursos de nube listos para ser utilizados |
| Seleccionar y configurar servicio SaaS para manejo de mapas. | Servicio Saas de mapas configurado |
| Desarrollo de microservicio BE integrador, para extraer la data del sitio seleccionado y llevarla a nuestro repositorio. | Microservicio que ingeste la información de la fuente seleccionada |
| Generación de script en terraform (CI/CD) | Script de terraform para la integración continua y despliegue continuo de toda la solución |
| Desarrollo de microservicio FE, integrado con servicio SaaS para manejo de mapas. | Microservicio que presente en un mapa la posibilidad de colocar 4 punteros y mapas de calor |
| Realizar pruebas de integración del sistema propuesto. | Deck de pruebas de integración ejecutado exitosamente |
| Validar el modelo propuesto, mediante la comparación de datos resultado de la predicción y los datos observados. | En la fase final, se validará el modelo desarrollado al aplicarlo a un caso de estudio real, para posteriormente: documentar exhaustivamente el proceso, los resultados y difundir los hallazgos a la comunidad científica. Es por ello que se usará una metodología cuantitativa ya que la idea es poder encontrar, para luego exponer el conocimiento encontrado | Recopilar datos observados de la eficiencia de paneles solares para el caso de estudio propuesto. | Los datos observados, en el horizonte cercano estarán almacenados en la DB |
| Realizar las conclusiones a las que haya lugar al respecto de la investigación | Documento con conclusiones relevantes del proyecto de investigación y sus evidencias |
| Realizar pruebas para el caso de estudio propuesto. | Reporte con pruebas del caso de estudio |
| Realizar paper con resultados finales de la investigación. | Paper con resultados finales |
| Realizar documento definitivo de la investigación. | Documento definitivo de la investigación |