



Pronóstico de la Radiación Solar para la Generación de Energía Eléctrica Fotovoltaica

Álvaro Jaramillo Duque

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Tutor

Walter Mauricio Villa Acevedo, Magíster (MSc)

Universidad de Antioquia

Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2022

Cita	Muñoz Zapata y Martínez Naranjo [1]
Referencia	[1] Á. Jaramillo Duque “Pronóstico de la Radiación Solar para la Generación de Energía Eléctrica Fotovoltaica”, Trabajo de grado especialización, Especialización en Analítica y Ciencia de Datos, Universidad de Antioquia, Medellín, Antioquia, Colombia, 2022.
Estilo IEEE (2020)	



Especialización en Analítica y Ciencia de Datos, Cohorte III.

Grupo de Manejo Eficiente de la Energía, GIMEL

Centro de Investigación Ambientales y de Ingeniería (CIA).



Centro de Documentación Ingeniería

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano/Director: Jesús Francisco Vargas Bonilla.

Jefe departamento: Diego José Luis Botía Valderrama.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

TABLA DE CONTENIDOS

1. RESUMEN EJECUTIVO	4
2. DESCRIPCIÓN DEL PROBLEMA	5
2.1 PROBLEMA DE NEGOCIO	5
2.2 APROXIMACIÓN DESDE LA ANALÍTICA DE DATOS	5
2.3 ORIGEN DE LOS DATOS	5
2.4 MÉTRICAS DE DESEMPEÑO	6
3. DATOS	8
3.1 DATOS ORIGINALES	8
3.2 DATASETS	9
3.3 DESCRIPTIVA	9
4. PROCESO DE ANALÍTICA	11
4.1 PIPELINE PRINCIPAL	11
4.2 PREPROCESAMIENTO	12
4.3 MODELOS	12
4.4 MÉTRICAS	14
5. METODOLOGÍA	15
5.1 HERRAMIENTAS	15
6. RESULTADOS	16
6.1 MÉTRICAS	16
6.2 EVALUACIÓN CUALITATIVA	17
6.3 CONSIDERACIONES DE PRODUCCIÓN	18
7. CONCLUSIONES	20
BIBLIOGRAFÍA	21

1. RESUMEN EJECUTIVO

El pronóstico de la energía solar fotovoltaica es cada vez más importante para los operadores de los sistemas eléctricos, dada la naturaleza incierta de la fuente primaria y su continua integración en estos. Por tanto, para una correcta operación de los sistemas eléctricos es necesario contar con pronósticos de la producción de la energía solar, lo más precisos posibles.

Este trabajo se enfocará en predicciones para un horizonte de tiempo de corto plazo (24 horas). Estos pronósticos tienen una aplicación directa en los mercados de energía eléctrica del tipo diario e intradiarios, para la estimación de reservas secundarias y terciarias, entre otros aspectos importantes de la operación de los sistemas eléctricos que incluyan fuentes de energía solar.

Actualmente, existe una tendencia hacia el uso de métodos de pronóstico basados en analítica de datos avanzada. Por lo tanto, el presente trabajo hace una comparación entre diferentes métodos de aprendizaje de máquina. La comparación consiste en entrenar y comparar diferentes modelos capaces de proporcionar series de tiempo de la radiación solar con un día de anticipación. Para estos experimentos se hace uso de un conjunto de datos abiertos proporcionado por el proyecto SolarMap. El conjunto de datos comprende diferentes variables climáticas y abarca un período que va desde el 6 de abril de 2010 hasta el 3 de abril de 2012.

Después de los experimentos realizados, se encontró que los modelos con redes neuronales Convolutional Neural Network combinados con Long short-term memory (CNN-LSTM) son los que lograron mejor desempeño comparadas con el resto de modelos implementados y probados.

Repositorio de GitHub [1]:

Á. Jaramillo-Duque, Solar Forecasting. 2022. Accedido: 25 de mayo de 2022. [En línea]. Disponible en: <https://github.com/alvarojd/SolarForecasting>

2. DESCRIPCIÓN DEL PROBLEMA

En los últimos años, la energía solar se ha convertido en una de las Fuentes No Convencionales de Energía Renovable (FNCR) más populares y se estima que en la próxima década será una de las fuentes más rentables por encima de las fuentes convencionales de energía [2]. Desafortunadamente, la generación de energía solar depende completamente del sol como recurso primario, el cual es un recurso incierto debido a los fenómenos atmosféricos en la tierra. Esta dependencia del clima agrega incertidumbre y variabilidad a la generación de energía solar fotovoltaica [3]. Para hacer frente a esta incertidumbre, se realizan pronósticos solares con el fin de predecir la generación futura de energía eléctrica fotovoltaica. Los pronósticos son necesarios para la correcta operación de los sistemas eléctricos [4].

2.1 PROBLEMA DE NEGOCIO

Para mitigar los problemas de la incertidumbre del recurso solar y los potenciales impactos en los sistemas eléctricos, es necesario contar con métodos de predicción lo más fiables posibles. Adicionalmente para reducir las emisiones de efecto invernadero es necesario utilizar al máximo los recursos energéticos renovables. Por tanto, más que un problema técnico es un compromiso de las empresas garantizar la cobertura de la demanda eléctrica con bajos niveles de gases de efecto invernadero [2]. Por lo anterior, la industria y la academia hacen esfuerzos para mejorar los métodos de pronóstico del recurso solar para garantizar la confiabilidad en el suministro de energía eléctrica a partir de las nuevas instalaciones fotovoltaicas.

2.2 APROXIMACIÓN DESDE LA ANALÍTICA DE DATOS

Este trabajo se enfoca en las predicciones a corto plazo (24 horas); estos pronósticos tienen una aplicación en los mercados de energía eléctrica del tipo diario e intradiarios, para la estimación de reservas secundarias y terciarias, entre otros aspectos de la operación de los sistemas eléctricos que incluyan fuentes de energía solar. Adicionalmente, las redes inteligentes son el nuevo paradigma de la operación de los sistemas eléctricos. Las redes inteligentes tienen como objetivo hacer un uso eficiente de los recursos energéticos y garantizar una correcta operación. Es por ello que es necesario desarrollar y contar con métodos de pronóstico de los recursos energéticos que tienen incertidumbre. De esta manera, se puede realizar la programación de las diferentes tecnologías de generación, en especial la que utilizan recursos energéticos intermitentes, y de almacenamiento para cubrir la demanda de energía eléctrica en un sistema eléctrico [5].

2.3 ORIGEN DE LOS DATOS

Dado que la solución del problema tiene un alcance académico, los datos no corresponden a una empresa o entidad. Por esta razón se buscaron datos genéricos y de origen abierto o público, que pudieran ser utilizados y que representarían las condiciones y necesidades de las empresas o instituciones del sector eléctrico.

Los datos constan de un conjunto de registros históricos relacionados con variables meteorológicas y geográficas tales como: irradiancias, temperatura del aire, humedad, velocidad del viento, ángulo cenital y ángulo de azimut, entre otras variables.

Con estas variables se espera estimar la irradiación del día siguiente, como una serie de tiempo basado en los datos del pasado.

2.4 MÉTRICAS DE DESEMPEÑO

Las métricas típicamente utilizadas en la literatura especializada y en la industria, para los pronósticos de la irradiancia y los recursos energéticos, son las siguientes:

Error absoluto medio (Mean Absolute Error o MAE)

En estadística, el error absoluto medio es una medida de la diferencia entre dos variables continuas. Por ejemplo, considerando dos series de datos o series de tiempo (unos estimados y otros observados), el error absoluto medio sirve para medir la precisión de la técnica de estimación usada, comparando los valores predichos frente a los observados.

Finalmente se obtiene un promedio de los errores. Si se obtiene un valor alto, representa un error alto del método de estimación.

Raíz del error cuadrático medio (Root Mean Square Error o RMSE)

Esta métrica consiste en calcular la raíz cuadrada del promedio de los errores cuadrados, es decir, se basa en las diferencias entre los valores predichos y los valores observados. Esta métrica sirve para agrupar las magnitudes de todos los errores de las predicciones, en una sola medida. El resultado es una medida de la precisión del modelo y su magnitud depende de la escala de los datos.

El RMSE siempre es positivo, y un valor de 0 indicaría una predicción perfecta. En general, un valor bajo siempre es mejor. Sin embargo, las comparaciones entre diferentes tipos de datos no son válidas ya que la métrica depende de la magnitud o la escala de valores analizados.

El efecto de cada error, en la medida final del RMSE, es proporcional al tamaño del error cuadrado. Por lo tanto, los errores más grandes, tienen un efecto desproporcionadamente grande en el valor final del RMSE. Por lo tanto, esta métrica es muy sensible a los valores atípicos.

Error Porcentual Absoluto Medio (Mean Absolute Percentage Error o MAPE)

Es una medida del desempeño del pronóstico que mide el tamaño del error absoluto, en términos porcentuales. Al ser una medida del error porcentual, lo hace un indicador frecuentemente utilizado por su fácil interpretación. Sin embargo las desviaciones en los valores de baja magnitud, inflan el error global y desviaciones de alta magnitud, subestiman el error global.

Métricas del pronóstico de la irradiancia.

Es complejo definir cuál es el error máximo admisible en la estimación del recurso solar, ya que depende de muchos factores y del sistema eléctrico al que esté conectado. Algunos de estos factores son la dimensión de la planta solar, compromisos de la energía generada, otras fuentes de energía o redes de suministro, sistemas de almacenamiento entre otros factores.

Según la literatura especializada [6], los errores reportados de la predicción de la irradiancia, en un intervalo de tiempo de 24 horas, son aproximadamente de un MAE: 5,8 y un RMSE: 11,9 con un buen modelo de estimación.

3. DATOS

3.1 DATOS ORIGINALES

Los datos de irradiación solar y variables climatológicas se obtuvieron del *National Renewable Energy Laboratory (NREL)*, proyecto SolarMap, Loyola Marymount University. disponibles en la página web: <https://midcdmz.nrel.gov/apps/day.pl?LMU> y con el requerimiento de citar la fuente para su uso [7].

El conjunto de datos comprende diferentes variables climáticas y abarca un período que va desde el 6 de abril de 2010 hasta el 3 de abril de 2012. Algunas de las variables que tiene el conjunto de datos son las siguientes:

- Irradiancia Global horizontal (GHI) [W/m^2]
- Irradiancia Normal Directa (DNI) [W/m^2]
- Irradiancia Difusa Horizontal (DHI) [W/m^2]
- Temperatura del aire [$^{\circ}\text{C}$]
- Humedad relativa [%]
- Velocidad del viento [m/s]
- Dirección del viento [$^{\circ}$ desde N]
- Presión atmosférica de la estación [mBar]
- Precipitación [mm]
- Ángulo cenital [$^{\circ}$]
- Ángulo de azimut [$^{\circ}$]

Estos valores originalmente tienen una resolución por minuto, por lo tanto, el conjunto de datos utilizado para el proceso de pronóstico comprende 1.048.576 registros.

Los datos se encuentran sin índice, que para este caso sería la fecha. Tampoco tienen encabezados, por lo que, después de la lectura del conjunto de datos, se procede a nombrar cada característica, correspondiente a todas las columnas del dataframe. Los encabezados de izquierda a derecha son los siguientes: 'Year', 'DOY', 'PST', 'GHI', 'DNI', 'DHI', 'Air Temp', 'Pressure', 'Wind Speed', 'Wind Dir', 'Wind Dir STD', 'Wind Speed STD', 'Wind Speed Peak', 'Global Uncorrected', 'Direct Uncorrected', 'Diffuse Uncorrected', 'Global Secondary', 'Unnamed: 18', 'Zenith', 'Azimuth', 'CR1000 Temp', 'Battery', 'Humidity', 'Precipitation', 'Precipitation Acc', 'Hour'.

El archivo original de texto plano tiene un tamaño total de 129 MB y se procesó para reducir la frecuencia a 15 minutos y también para eliminar datos atípicos o faltantes.

3.2 DATASETS

El conjunto de datos originales se procesó para reducir la frecuencia, pasando a registros cada quince minutos. También se descartaron algunas características, y solo se utilizaron para el proceso de entrenamiento y validación las siguientes:

- Irradiancia Global horizontal (GHI) [W/m^2]
- Día del año [día]
- Presión atmosférica de la estación [mBar]
- Precipitación [mm]
- Hora del día [h]
- Ángulo cenital [$^\circ$]
- Humedad relativa [%]

Este nuevo conjunto de datos se almacenó en el repositorio de GitHub [1] dado que no supera las 6MB y se cargan directamente desde los notebook.

Se separaron los datos de la serie de tiempo para entrenamiento y validación, que corresponden a un 80% y 20% respectivamente.

También se revisaron otros conjuntos de datos y están disponibles en el repositorio de GitHub [1]. Estos conjuntos de datos son semejantes al anterior mencionado y se obtuvieron de la siguiente fuente [8].

3.3 DESCRIPTIVA

Las estadísticas descriptivas de los datos procesados, incluyen la tendencia central, la dispersión y entre otras características del conjunto de datos, se presentan en la Tabla I. Todos los detalles del procesamiento y limpieza de los datos, están en el repositorio en el notebook llamado 00-Carga y limpieza de datos.ipynb.

TABLA I
ESTADÍSTICAS DESCRIPTIVAS DE LOS DATOS PROCESADOS DEL CONJUNTO DE DATOS [7]

	GHI	DOY	PST	Precipitation	Zenith	Humidity
count	69845,00	69845,00	69845,00	69845,00	69845,00	69845,00
mean	238,21	183,30	1172,09	0,00	90,79	75,76
std	331,53	105,42	692,40	0,03	39,09	19,00
min	0,00	1,00	0,00	0,00	10,78	7,48
25%	0,00	91,00	545,00	0,00	58,59	67,05
50%	5,96	184,00	1145,00	0,00	90,80	80,20
75%	462,87	275,00	1745,00	0,00	122,60	90,30
max	1494,38	365,00	2345,00	1,52	169,50	100,00

La irradiancia de un día típico se puede ver en la Fig. I. En este caso se puede evidenciar un día despejado donde no hay reducción de la radiación por eventos atmosféricos o nubes.

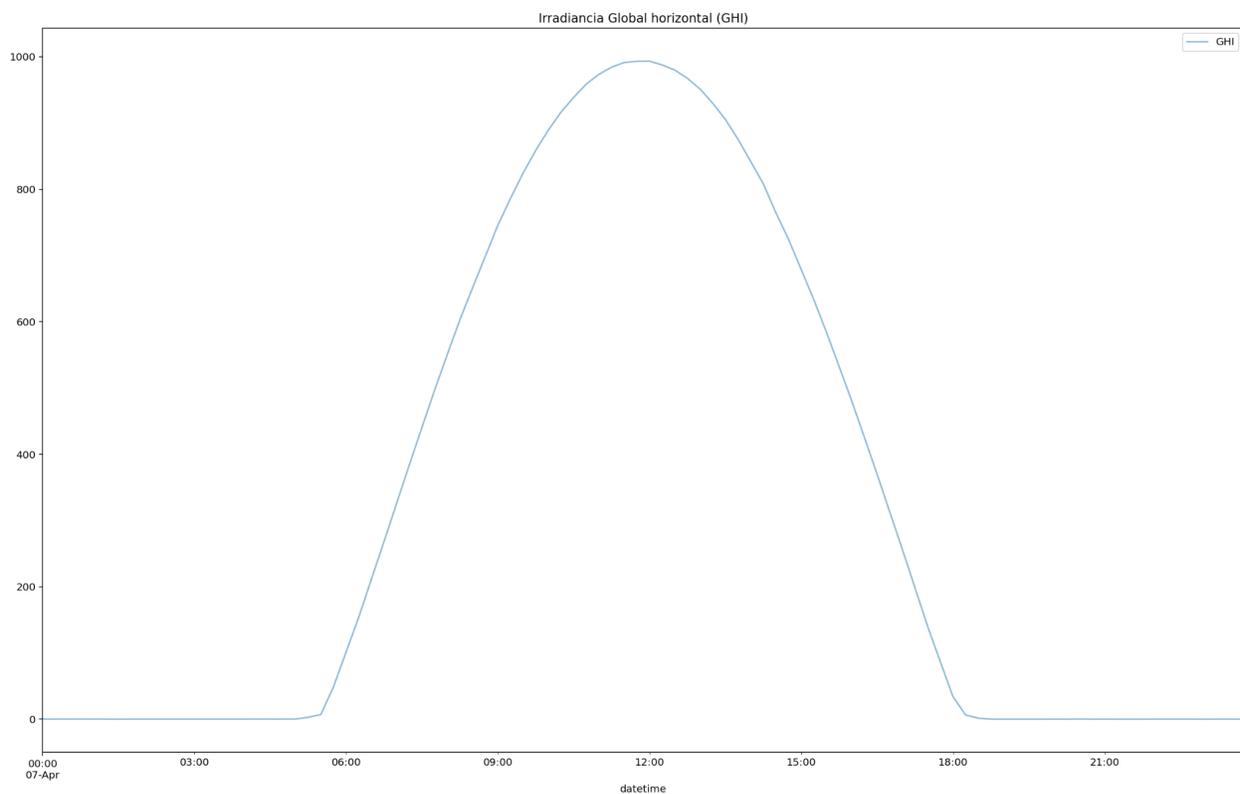


Fig. I: Irradiancia global horizontal en $[W/m^2]$, para un día del conjunto de datos [7].

4. PROCESO DE ANALÍTICA

4.1 PIPELINE PRINCIPAL

En la Fig. II se muestra el diagrama de flujo que representa el procedimiento desde que se procesan los datos, luego la construcción del modelo, entrenamiento y validación. Finalmente se calculan las métricas y se verifica el funcionamiento del modelo graficando algunos de los resultados.

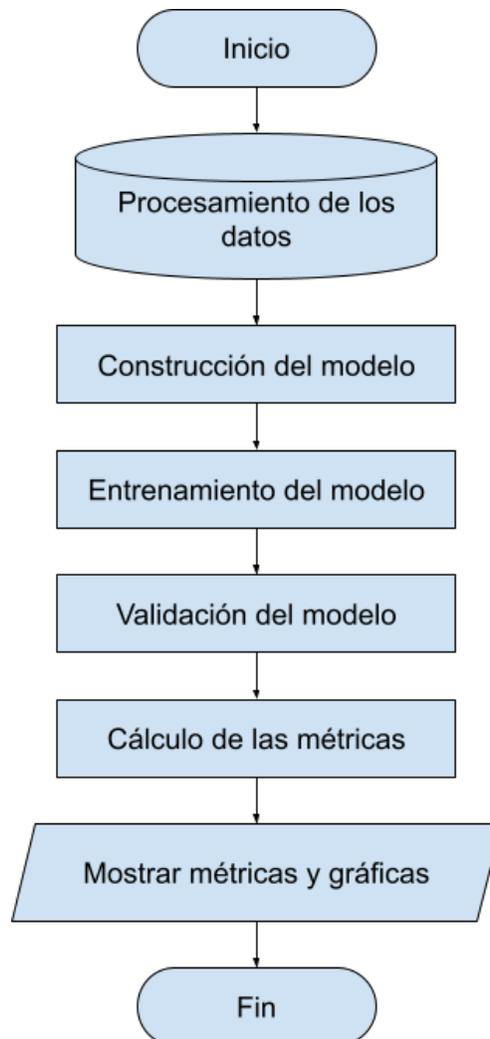


Fig. II: Diagrama de flujo que representa el flujo de trabajo para la solución del problema .

4.2 PREPROCESAMIENTO

Los datos originalmente tienen una resolución por minuto. El conjunto de datos utilizado para el proceso de pronóstico comprende 1.048.576 registros. El conjunto de datos originales se procesó para reducir la frecuencia, pasando a registros cada quince minutos, llegando a un total de 69.906 filas de valores. Para este caso se utilizó el valor máximo para evitar problemas con datos faltantes.

Posteriormente se escribió una función llamada “ventanas”, para organizar la información y procesar datos por días. Ya que el objetivo es estimar la irradiancia para el día siguiente, partiendo de los datos del día pasado.

Posteriormente se separaron los datos de la serie de tiempo ajustada por días, para entrenamiento y validación, que corresponden a un 80% y 20% respectivamente.

Finalmente se construyen dos *data frames* con las dimensiones ajustadas para realizar el entrenamiento y validación de los modelos. Estas dimensiones corresponden a la forma como se le deben ingresar los datos a las redes LSTM y CNN.

Este mismo proceso se realizó con los datos obtenidos de la fuente alternativa de datos [8].

4.3 MODELOS

Las redes neuronales artificiales (Artificial Neural Network o ANN), son una familia de algoritmos que conforman la base de la denominada inteligencia artificial, a la cual se le ha llamado aprendizaje profundo (Deep Learning o DL). Este tipo de técnicas de inteligencia artificial han tenido gran acogida por presentar resultados satisfactorios en la predicción de comportamientos de variables o recursos de interés, entre estas las series de tiempo.

Como objetivo se planteó una comparación de diferentes modelos de deep learning, entre ellos se consideraron las siguientes redes neuronales y algunas combinaciones entre ellas:

- Long short-term memory (LSTM)
- Convolutional Neural Network (CNN)
- Gated Recurrent Unit (GRU)

Las ANN son un conjunto de funciones y operaciones matemáticas que procesan un conjunto de datos y generan una salida. A continuación se hará una definición de estos conceptos.

Red Neuronal Artificial (Artificial Neural Network o ANN)

Consiste en un conjunto de neuronas artificiales que están conectadas entre sí con el objetivo de transmitir las señales de entrada y salida. Este concepto partió del interés en entender y simular una neurona real. Cada valor de la entrada de una neurona es multiplicado por un peso, luego todas las entradas son sumadas y pasan a la salida de la neurona donde se le aplica una función de activación. Esta se encarga de

modificar o limitar el valor de salida que será propagado a las otras neuronas de las siguientes capas. Los valores asignados a los pesos, en la entrada de las neuronas, son ajustados por medio de la minimización de una función de pérdida, a este proceso se le llama retro propagación hacia atrás y se realiza en la etapa de entrenamiento de la red. Una vez la red está entrenada, se puede utilizar para realizar una tarea [9].

Red Neuronal Recurrente (Recurrent Neural Network o RNN):

Este tipo de redes presenta una mayor capacidad para entender y aprender diferentes patrones en los datos de entrada. Las RNN no solo incluyen conexiones hacia adelante como las redes neuronales convencionales, sino también conexiones hacia atrás, creando un tipo de retroalimentación. Esto le permite exhibir un comportamiento dinámico temporal [9].

Gran Memoria a corto plazo (Long Short-Term Memory o LSTM)

Las redes LSTM son una versión de las redes RNN tratadas anteriormente, que pretende mejorar el desempeño al evitar problemas de convergencia en el proceso de entrenamiento. Las redes LSTM tienen como característica principal que la información puede persistir introduciendo bucles en el grafo de la red, por lo que pueden recordar estados previos y utilizar esta información para decidir cuál será el estado siguiente. Esta característica las hace muy adecuadas para estimar series temporales [9].

Unidad Recurrente Cerrada (Gated Recurrent Units GRU)

Este tipo de redes parte del concepto de una RNN, que pretende usar conexiones a través de una secuencia de nodos para realizar tareas de aprendizaje automático asociadas con la memoria y la agrupación en clústeres. Este tipo de redes presentan una puerta de actualización y una puerta de reinicio, con las cuales el modelo controla la información y la puede retener durante un periodo de tiempo [9].

Red Neuronal Convolutiva (Convolutional Neural Network o CNN)

Las CNN son versiones regularizadas de perceptrones multicapa. Los perceptrones multicapa generalmente significan redes completamente conectadas, es decir, cada neurona en una capa está conectada a todas las neuronas en la siguiente capa [9].

Salida de los modelos y otras características

También se consideraron dos tipos de modelos para las redes, univariante y multivariante. Para ambos casos se estimó el siguiente día. En todos los casos se aplicó el concepto de secuencia a vector, donde el vector de salida contiene los datos de un día completo.

Además se consideraron dos tipos de normalización de los datos, de la librería scikit-learn, StandardScaler y MinMaxScaler [10].

Como aspectos generales para todos los modelos implementados, se consideraron las siguientes características:

- Dos capas, la primera de 100 neuronas y la segunda de 150 neuronas.
- Función de activación tangente hiperbólica
- Optimizador Adam
- Función de pérdidas MAE

- Cien épocas de entrenamiento máximo
- Se utilizó la función Dropout con un valor del 20%.

Finalmente se realizaron los siguientes modelos:

- LSTM_Univariable_Standard
- LSTM_Univariable_MinMax
- LSTM_Multivariable_Standard
- LSTM_Multivariable_MinMax
- CNN_LSTM_Univariable_Standard
- CNN_LSTM_Univariable_MinMax
- CNN_LSTM_Multivariable_Standard
- CNN_LSTM_Multivariable_MinMax
- GRU_Univariable_Standard
- GRU_Univariable_MinMax
- GRU_Multivariable_Standard
- GRU_Multivariable_MinMax
- CNN_GRU_Univariable_Standard
- CNN_GRU_Univariable_MinMax
- CNN_GRU_Multivariable_Standard
- CNN_GRU_Multivariable_MinMax

En el repositorio de GitHub se encuentran organizados por carpetas los diferentes modelos. Hay un notebook por cada modelo, el cual es ejecutable desde google Colab y lee los datos de forma automática.

4.4 MÉTRICAS

Se calcularon las métricas MAE, MAPE y RMSE para cada uno de los modelos, tanto para el conjunto de entrenamiento y validación. Para estos cálculos se implementó un código propio y posteriormente se tabularon los datos para poder comparar los diferentes modelos.

5. METODOLOGÍA

Dado que el objetivo era comparar diferentes modelos, se partió de modelos simples a modelos más sofisticados o con tipos de redes que pudieran tener un mejor desempeño para pronosticar series de tiempo irradiancia.

Los modelos comparten varias de las características con el objetivo de poder comprobar su desempeño, tal como se explicó en la sección 4.3. Los datos de entrenamiento y validación fueron los mismos para todos los modelos. Es importante mencionar que como es un problema de pronóstico los dos conjuntos (entrenamiento y validación) deben conservar sus características temporales es decir, el conjunto de prueba deben ser los datos más actuales de la serie de tiempo a pronosticar. Igualmente en el entrenamiento las observaciones de series de tiempo de mantener su orden temporal, es decir, no se realiza “shuffle” en los datos del conjunto

Para validar la robustez de los diferentes modelos implementados, se utilizaron dos conjuntos de datos. Más adelante se presentan los resultados de los mismos y se comparan las métricas para identificar el modelo que consigue estimar de forma más exacta la irradiación del día siguiente.

5.1 HERRAMIENTAS

Para el desarrollo e implementación de los diferentes modelos se utilizó la librería Tensor Flow, Pandas, scikit-learn y otras librerías ampliamente utilizadas en Python para la lectura, procesamiento y visualización de los datos. Todos los detalles se encuentran en los notebooks en el repositorio GitHub [1].

6. RESULTADOS

6.1 MÉTRICAS

En la Tabla II están registrados los resultados obtenidos para los dieciséis diferentes modelos implementados para este estudio. Se marcan en negrita los cuatro mejores resultados en los datos de validación (errores más bajos). Dado estos resultados, se puede determinar que las redes neuronales Convolutional Neural Network combinadas con redes Long short-term memory (CNN_LSTM) obtienen tres de las mejores métricas y la red Gated Recurrent Unit ocupa el cuarto lugar.

TABLA II
RESULTADOS DEL PRIMER CONJUNTO DE DATOS [7]

Nombre del modelo	Entrenamiento [W/m ²]		Validación [W/m ²]	
	MAE	RMSE	MAE	RMSE
LSTM_Univariable_Standard	43,94	74,89	42,53	75,51
LSTM_Univariable_MinMax	56,96	94,17	46,20	82,18
LSTM_Multivariable_Standard	41,33	71,54	51,14	88,79
LSTM_Multivariable_MinMax	46,58	82,61	40,84	73,74
CNN_LSTM_Univariable_Standard	38,39	66,41	38,86	70,43
CNN_LSTM_Univariable_MinMax	53,94	90,99	40,43	73,98
CNN_LSTM_Multivariable_Standard	34,30	63,33	37,37	67,45
CNN_LSTM_Multivariable_MinMax	48,20	85,73	42,66	76,42
GRU_Univariable_Standard	91,46	131,36	96,51	142,07
GRU_Univariable_MinMax	148,39	201,25	130,60	182,47
GRU_Multivariable_Standard	68,12	102,14	56,83	87,87
GRU_Multivariable_MinMax	49,15	85,01	39,73	71,58
CNN_GRU_Univariable_Standard	106,04	163,66	61,39	99,30
CNN_GRU_Univariable_MinMax	136,65	178,46	150,59	197,23
CNN_GRU_Multivariable_Standard	63,18	95,20	59,28	92,03
CNN_GRU_Multivariable_MinMax	50,96	85,70	46,84	80,31

En la Tabla III están registrados los resultados obtenidos para los dieciséis diferentes modelos implementados para este estudio. Se marcan en negrita los cuatro mejores resultados en los datos de validación (errores más bajos). Dado estos resultados, no se puede determinar cuál tipo de modelo obtiene el mejor desempeño. Sin embargo, se puede apreciar que las redes LSTM, CNN y GRU, obtienen los mejores resultados.

Tabla III
Resultados del segundo conjunto de datos [8]

Nombre del modelo	Entrenamiento [W/m ²]		Validación [W/m ²]	
	MAE	RMSE	MAE	RMSE
LSTM_Univariable_Standard	40,42	78,57	40,55	77,42
LSTM_Univariable_MinMax	50,21	93,92	38,99	74,92
LSTM_Multivariable_Standard	47,89	89,49	39,46	74,88
LSTM_Multivariable_MinMax	49,83	94,32	39,94	76,53
CNN_LSTM_Univariable_Standard	46,38	87,50	41,05	77,07
CNN_LSTM_Univariable_MinMax	51,38	95,87	39,15	75,47
CNN_LSTM_Multivariable_Standard	67,58	117,72	40,17	76,94
CNN_LSTM_Multivariable_MinMax	49,88	93,31	39,25	75,12
GRU_Univariable_Standard	72,15	121,65	43,79	77,05
GRU_Univariable_MinMax	76,69	121,12	52,97	83,78
GRU_Multivariable_Standard	72,82	124,84	41,13	76,47
GRU_Multivariable_MinMax	49,67	92,78	38,64	74,34
CNN_GRU_Univariable_Standard	77,17	127,89	46,48	80,88
CNN_GRU_Univariable_MinMax	85,07	121,12	75,57	104,07
CNN_GRU_Multivariable_Standard	75,68	126,18	44,06	78,39
CNN_GRU_Multivariable_MinMax	48,84	91,96	39,08	74,50

De las tablas anteriores, se puede evidenciar que estos tipos de modelos, llegan a obtener buenos resultados en general. Además que son modelos de baja complejidad. Para poder llegar a una conclusión más general, sería necesario ejecutar los mismos experimentos con conjuntos de datos adicionales. De esta forma se podría generalizar el tiempo de modelo que mejor estima este tipo de serie de tiempo.

6.2 EVALUACIÓN CUALITATIVA

Las predicciones se realizaron para todo el conjunto de validación, sin embargo, para mostrar el comportamiento del modelo se seleccionaron de forma aleatoria algunos días.

En la Fig. III y Fig. IV, se puede observar el resultado de la predicción de dos días de los datos validación del segundo conjunto de datos [8]. En las figuras, el eje horizontal corresponde al tiempo, datos cada 15 minutos; el eje vertical corresponde a la irradiancia en [W/m²]. La línea azul es la irradiancia del último día; la línea roja punteada, es la estimación del día siguiente y la línea verde punteada, es la observación de la irradiancia. Tal como se puede ver en ambas figuras, la línea roja sigue la tendencia del comportamiento de la línea verde. Sin embargo, hay algunos momentos donde no se logra seguir la tendencia, esto se debe a que es difícil estimar los fenómenos atmosféricos y la cantidad de nubes que afectan la irradiancia.

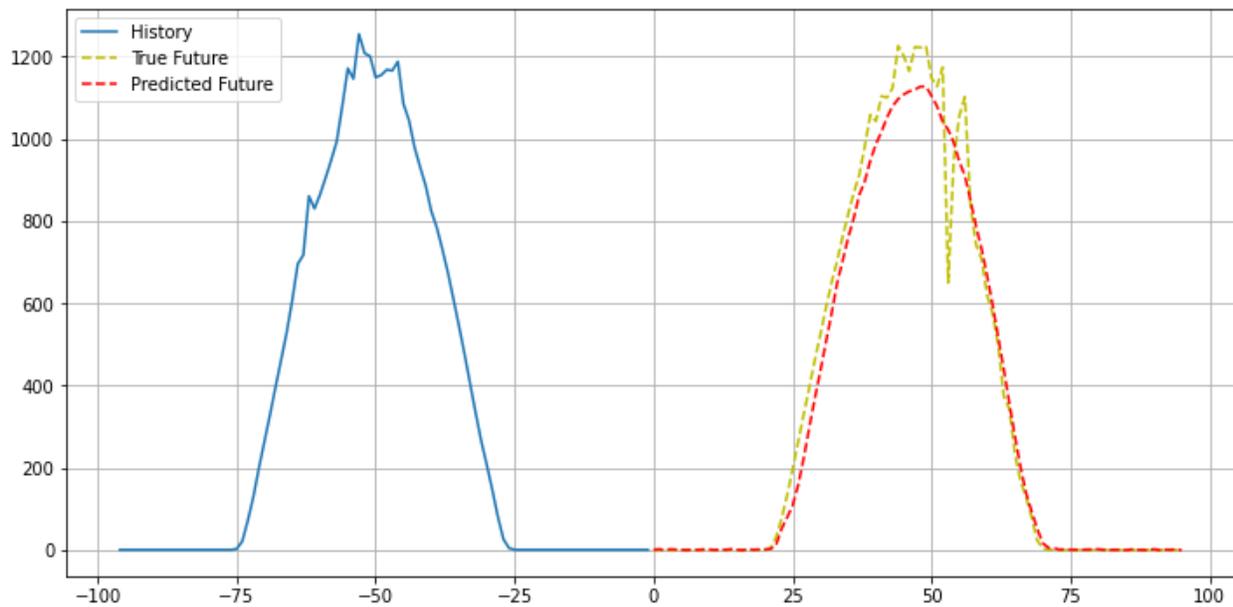


Fig. III: Estimación de la irradiación con el modelo CNN_LSTM_multivariable_Array_Standard.

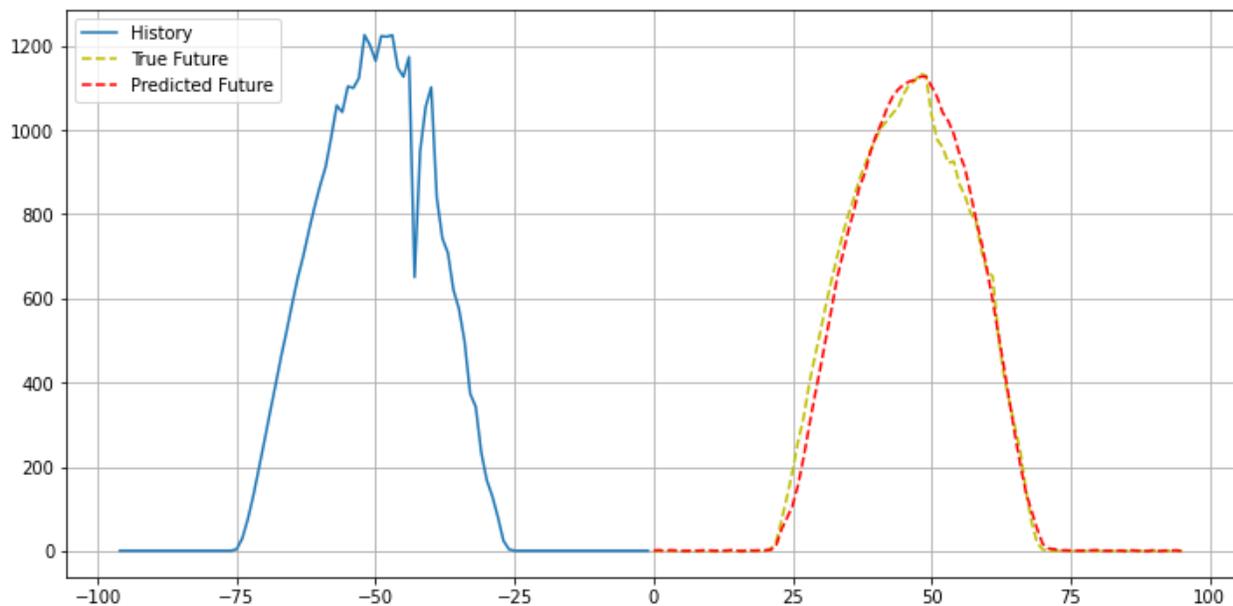


Fig. IV: Estimación de la irradiación con el modelo CNN_LSTM_multivariable_Array_Standard.

6.3 CONSIDERACIONES DE PRODUCCIÓN

Este tipo de modelos tienen una gran facilidad de llevarse a producción ya que los procesos de entrenamiento y validación pueden ser ejecutados con una periodicidad de una semana. Mientras que los tiempos de estimación son muy cortos, ya que solo se predice el día siguiente, cada día.

Es posible crear modelos más sofisticados y que puedan contener variables o características adicionales para estimar los atmosféricos y la cantidad de nubes que afectan la irradiancia. De esta manera se podría mejorar la tendencia de las predicciones realizadas.

7. CONCLUSIONES

Todos los modelos de redes neuronales resultan una gran herramienta para la predicción de la irradiancia solar si se realiza un tratamiento adecuado a los datos y si se escogen parámetros adecuados. Sin embargo, es necesario hacer más experimentos para determinar la arquitectura de la red y una selección de parámetros para poder definir cuál es el mejor modelo, esto se logra implementando un proceso de sintonización de hiper parámetros de las redes neuronales profundas.

Los modelos con redes neuronales, permiten hacer predicciones con modelos univariantes, donde solo se considera la irradiación y modelos multivariantes, donde se consideran otras características que pueden ayudar a estimar la irradiación. En ambos casos se lograron buenos resultados, lo que puede significar que una combinación de estos modelos puede conseguir un buen resultado.

BIBLIOGRAFÍA

- [1] Á. Jaramillo-Duque, *Solar Forecasting*. 2022. Accedido: 25 de mayo de 2022. [En línea]. Disponible en: <https://github.com/alvarojd/SolarForecasting>
- [2] REN21, «Renewables Global Status Report», *REN21*. <https://www.ren21.net/reports/global-status-report/> (accedido 2 de abril de 2022).
- [3] J. Urrego-Ortiz, J. A. Martínez, P. A. Arias, y Á. Jaramillo-Duque, «Assessment and Day-Ahead Forecasting of Hourly Solar Radiation in Medellín, Colombia», *Energies*, vol. 12, n.º 22, Art. n.º 22, ene. 2019, doi: 10.3390/en12224402.
- [4] J. E. Sierra-Aguilar, C. C. Marín-Cano, J. M. López-Lezama, Á. Jaramillo-Duque, y J. G. Villegas, «A New Affinely Adjustable Robust Model for Security Constrained Unit Commitment under Uncertainty», *Appl. Sci.*, vol. 11, n.º 9, Art. n.º 9, ene. 2021, doi: 10.3390/app11093987.
- [5] G. Gürses-Tran y A. Monti, «Advances in Time Series Forecasting Development for Power Systems' Operation with MLOps», *Forecasting*, vol. 4, n.º 2, Art. n.º 2, jun. 2022, doi: 10.3390/forecast4020028.
- [6] M. Abdel-Nasser, K. Mahmoud, y M. Lehtonen, «HIFA: Promising Heterogeneous Solar Irradiance Forecasting Approach Based on Kernel Mapping», *IEEE Access*, vol. 9, pp. 144906-144915, 2021, doi: 10.1109/ACCESS.2021.3122826.
- [7] T. Stoffel y A. Andreas, «Solar Resource & Meteorological Assessment Project (SOLRMAP): Rotating Shadowband Radiometer (RSR); Los Angeles, California (Data)», National Renewable Energy Lab. (NREL), Golden, CO (United States), NREL/DA-5500-56502, abr. 2010. doi: 10.7799/1052230.
- [8] ENERGY DATA, «Datasets - ENERGYDATA.INFO». https://energydata.info/dataset?res_format%3DCSV%26vocab_topics%3DSolar (accedido 15 de mayo de 2022).
- [9] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Beijing China ; Sebastopol, CA, 2019.
- [10] Raghav RV, Guillaume Lemaitre, y Thomas Unterthiner, «Compare the effect of different scalers on data with outliers», *scikit-learn*. https://scikit-learn/stable/auto_examples/preprocessing/plot_all_scaling.html (accedido 25 de marzo de 2022).