

Reporte Final de Proyecto

Grupo 3: Forecasting de ERNC

Ignacio Arriagada, José Audala y Joaquín Ormazábal A.

Departamento de Ingeniería Eléctrica, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile

14 de Diciembre, 2024

Abstract—Este proyecto aborda la predicción de la generación de energía solar mediante técnicas avanzadas de forecasting, con el objetivo de optimizar la integración de las Energías Renovables No Convencionales (ERNC) en la red eléctrica. Utilizando datos históricos y modelos estadísticos y de machine learning, se busca predecir la producción de energía solar a corto plazo, aplicando así conceptos vistos durante el curso, como: estructuras de datos y programación orientada a objetos. Los resultados obtenidos permitirán prever la variabilidad de la generación solar en la planta FV escogida, y facilitarán la toma de decisiones operativas en tiempo real, contribuyendo a una mayor estabilidad y sostenibilidad del sistema eléctrico.

I. INTRODUCTION

Las Energías Renovables No Convencionales (ERNC), particularmente la energía solar, se han consolidado como una de las principales alternativas en la transición hacia un modelo energético más sostenible. El aumento de la conciencia global sobre los efectos negativos del uso de combustibles fósiles, junto con la necesidad urgente de mitigar los impactos del cambio climático, ha impulsado el desarrollo y la implementación de tecnologías de generación de energía renovable. En este contexto, la energía solar fotovoltaica destaca como una de las fuentes más prometedoras debido a su accesibilidad, menor impacto ambiental y capacidad de generar electricidad de manera eficiente en una amplia gama de ubicaciones geográficas.

A pesar de sus ventajas, la energía solar presenta desafíos significativos en cuanto a su integración y gestión dentro de los sistemas eléctricos tradicionales. Estos desafíos están principalmente asociados con la intermitencia e incertidumbre inherentes a la generación solar. A diferencia de las fuentes de energía convencionales, como las térmicas o las hidroeléctricas, la energía solar depende de factores meteorológicos variables y no controlables, como la radiación solar, la nubosidad, la posición del sol y las condiciones climáticas locales, lo que hace que la generación de electricidad no sea constante y esté sujeta a fluctuaciones.

Por lo tanto, es esencial contar con métodos de pronóstico o forecasting que permitan predecir la producción de energía solar con el mayor grado de precisión posible. Estos métodos ayudan a reducir la incertidumbre, mejorando la planificación operativa y el manejo de la oferta y demanda energética en tiempo real. Las técnicas de forecasting de generación solar se han convertido en una herramienta indispensable para las empresas de distribución eléctrica, los operadores

de plantas solares y los responsables de la planificación de redes, ya que permiten optimizar la asignación de recursos, mejorar la eficiencia en la integración de la energía renovable y minimizar los costos asociados con el desajuste entre la oferta y la demanda energética.

El forecasting de la generación solar no solo es relevante para la gestión de la energía, sino que también tiene implicaciones profundas para la gestión de redes inteligentes y la optimización de almacenamiento de energía. A medida que las redes eléctricas avanzan hacia un modelo más flexible y descentralizado, basado en la integración de fuentes de energía renovables, el pronóstico de la producción solar se convierte en una herramienta crítica para equilibrar la oferta energética y garantizar la estabilidad del sistema.

Este proyecto tiene como objetivo desarrollar y evaluar un modelo de forecasting para la generación de energía solar, utilizando herramientas de análisis de datos y técnicas de machine learning para predecir la producción a corto plazo. Para ello, en este informe se analizarán diferentes enfoques para el modelado de la variabilidad de la radiación solar, con el fin de identificar las mejores metodologías para mejorar la precisión de las predicciones. En particular, se explorarán los métodos basados en redes neuronales y modelos estadísticos avanzados, que han demostrado ser eficaces para predecir fenómenos de series temporales no lineales y estocásticos.

II. DESCRIPCIÓN DEL PROBLEMA

La generación de energía solar fotovoltaica está influenciada por un conjunto de factores que varían en el tiempo y el espacio, lo que introduce un nivel significativo de incertidumbre. La radiación solar, principal fuente de energía para los paneles solares, está determinada por condiciones atmosféricas como la nubosidad, la temperatura, la humedad, y la hora del día, entre otros factores. A su vez, estas variables presentan un comportamiento altamente no lineal y estocástico, lo que dificulta su modelado y predicción. La interacción de estos factores y su variabilidad temporal hacen que la generación solar sea inherentemente intermitente, con fluctuaciones que no pueden preverse con exactitud a largo plazo.

Uno de los principales retos en la integración de la energía solar en las redes eléctricas es la sincronización

entre la generación y el consumo de electricidad. Las redes tradicionales están diseñadas para operar con fuentes de energía convencionales, que suelen tener una producción más predecible y controlada. En cambio, las fuentes solares producen electricidad de forma variable, lo que obliga a los operadores de red a ajustar constantemente la oferta de energía de otras fuentes o recurrir al almacenamiento de electricidad, lo que puede resultar costoso y técnicamente complejo.

En este sentido, el forecasting se presenta como una solución esencial para mitigar los efectos de la variabilidad solar. El objetivo de un modelo de forecasting es predecir con la mayor precisión posible la cantidad de energía que se generará a partir de los paneles solares en intervalos de tiempo a corto y medio plazo (horas, días). Sin embargo, existen múltiples dificultades en la predicción de la generación solar, especialmente debido a los siguientes factores:

- **Condiciones Climáticas Locales:** Las condiciones meteorológicas locales, como las nubes, la humedad y la temperatura, afectan directamente la radiación solar. La nubosidad es uno de los factores más complejos de predecir, ya que puede cubrir o descubrir áreas de generación solar en un corto período de tiempo, creando una variabilidad considerable en la producción.
- **Estacionalidad:** La radiación solar varía no solo durante el día, sino también de acuerdo con la estación del año. Las estaciones más soleadas, como la primavera y el verano, presentan una mayor producción solar, mientras que en invierno, la radiación solar disminuye debido a la inclinación de la Tierra y las horas de luz reducidas.
- **Incertidumbre en el Comportamiento de las Series Temporales:** La variabilidad inherente a la generación solar introduce ruido estocástico en los datos. Esto hace que los modelos de predicción tradicionales, basados en suposiciones de linealidad, no sean adecuados para capturar la complejidad del fenómeno. En cambio, los modelos basados en redes neuronales y algoritmos de machine learning son más efectivos para modelar la naturaleza no lineal de los datos y las relaciones entre los factores que afectan la radiación solar.
- **Desajuste entre Oferta y Demanda:** En muchos casos, la producción solar no coincide con los picos de demanda energética, lo que puede generar exceso de energía o déficit en la red. Si la predicción es inexacta, puede ocurrir un desajuste entre la generación prevista y la demanda real, lo que obliga a recurrir a fuentes de energía convencionales de respaldo o a aumentar la capacidad de almacenamiento. Esto puede implicar costos adicionales y una menor eficiencia en el uso de los recursos renovables.
- **Impacto en la Gestión de la Red Eléctrica:** Las predicciones imprecisas afectan la capacidad de los operadores de red para equilibrar la oferta y la demanda de electricidad. En sistemas eléctricos con alta penetración de energía solar, las fluctuaciones inesperadas pueden causar sobrecargas o fallas en la red si no se gestionan adecuadamente.

Dado este contexto, es crucial contar con modelos predictivos avanzados que no solo proporcionen estimaciones precisas de la generación solar a corto plazo, sino que también consideren la incertidumbre inherente a estos fenómenos. La capacidad de anticipar la producción solar con precisión es esencial para optimizar las decisiones operativas en tiempo real, como la gestión de la carga de la red, la programación de la generación de energía convencional y el almacenamiento de energía. Además, la mejora en los métodos de forecasting puede facilitar el desarrollo de redes eléctricas inteligentes y contribuir significativamente a la descarbonización del sector energético.

III. BIBLIOGRAFÍA CONSULTADA

En búsqueda de un modelo ML apropiado para el desarrollo de este proyecto, se revisó una serie de artículos y papers que proponen modelos para la implementación de *forecasting* en diversos contextos (no solo el atinente a este proyecto). En particular, se revisó el artículo [3] sobre predicción de potencia solar fotovoltaica utilizando un modelo LSTM con tasa de aprendizaje adaptativa *Adam* y tamaño de ventana optimizado. A continuación, se detalla más sobre éste.

El estudio propone un modelo de predicción de energía solar basado en Long Short-Term Memory (LSTM) con dos innovaciones principales: tasa de aprendizaje adaptativa y tamaño de ventana optimizado.

El modelo utiliza datos multivariados que incluyen:

- Irradiancia directa
- Irradiancia difusa
- Temperatura ambiente

La arquitectura del modelo LSTM se compone de las siguientes capas:

Entrada \rightarrow LSTM (64 unidades) \rightarrow Capa Densa ($8n_h$) \rightarrow Salida ($1n_h$)

donde se denota n_h el número de neuronas en la capa oculta h .

Los autores implementaron una estrategia de selección de hiperparámetros:

- Tasa de aprendizaje: Se evaluaron múltiples valores, seleccionando 0.02 como óptimo
- Tamaño de ventana: Se probaron valores entre 5 y 25, identificando 14 como el más efectivo

La performance del modelo se evaluó utilizando el Error Cuadrático Medio (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{\text{pred}} - x_{\text{real}})^2} \quad (1)$$

Los resultados mostraron una reducción significativa del RMSE:

Las contribuciones fundamentales del estudio incluyen:

- Integración de tasa de aprendizaje adaptativa
- Optimización del tamaño de ventana para capturar dependencias temporales

Método	RMSE
LSTM Simple	0.185
Bi-LSTM	0.248
LSTM-Transformer	1.691
CNN-LSTM	1.018
LSTM Propuesto	0.153

- Mejora de la precisión en predicción de generación solar

Los autores reconocen limitaciones en su enfoque:

- Enfoque en predicción de un solo paso
- Necesidad de explorar predicciones a múltiples pasos
- Potencial mejora mediante:
 - Métodos ensemble
 - Modelos híbridos
 - Incorporación de datos meteorológicos en tiempo real

En conclusión, la estructura LSTM propuesta por Kunalan [3]. representa una opción sumamente atractiva para proyectos de forecasting debido a su capacidad única para capturar dependencias temporales complejas. A diferencia de modelos tradicionales que tratan los datos de series temporales como observaciones independientes, los LSTM incorporan mecanismos de memoria que les permiten retener y procesar información relevante a lo largo de secuencias temporales extensas. En contextos de predicción de energía, donde las variables meteorológicas y de generación presentan correlaciones no lineales y dinámicas complejas, la arquitectura con compuertas (input, forget y output gates) resulta especialmente ventajosa. La posibilidad de implementar tasas de aprendizaje adaptativas y optimizar el tamaño de ventana añade capas adicionales de flexibilidad, permitiendo que el modelo se ajuste de manera más precisa a los patrones intrínsecos de los datos. Particularmente en sistemas de energía renovable, donde la variabilidad y la dependencia temporal son críticas, un modelo LSTM puede capturar fluctuaciones de corto plazo, tendencias estacionales y efectos de memoria de manera significativamente más efectiva que modelos estadísticos tradicionales.

IV. PROPUESTA Y JUSTIFICACIÓN

En inspiración del artículo revisado en la sección anterior [3], se propone utilizar una red LSTM como modelo de predicción de generación solar. Esta red tiene una estructura interna que le permite recordar información durante períodos de tiempo mucho más largos que las redes neuronales recurrentes convencionales [1]. Esto se logra mediante el uso de unidades de memoria especializadas y puertas de control. Las LSTM tienen una estructura con tres puertas principales, tal como se observa en la figura [1].

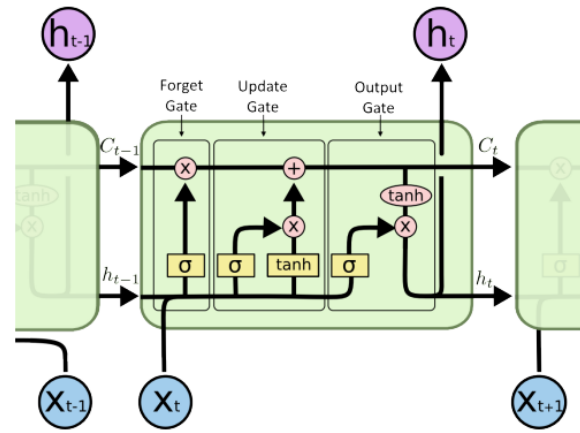


Fig. 1: Diagrama de una célula de memoria LSTM

1) Puerta de Olvido (Forget Gate):

- Esta puerta decide qué información se descarta de la celda de memoria.
- Toma como entrada el valor de la celda de memoria anterior y la entrada actual.
- La salida de esta puerta es un valor entre 0 y 1, donde 0 significa "olvidar" completamente y 1 significa "mantener" completamente la información.

2) Puerta de Entrada (Input Gate):

- Esta puerta decide qué información nueva se agrega a la celda de memoria.
- Se compone de dos partes: una que decide qué valores actualizar (usando una función sigmoide), y otra que crea un vector de nuevas celdas de memoria a través de una función tangente hiperbólica (tanh).
- El valor final agregado a la memoria es una combinación de los valores anteriores y la nueva información.

3) Puerta de Salida (Output Gate):

- Esta puerta decide qué parte de la celda de memoria se va a utilizar para la salida de la red.
- En primer lugar, se pasa por una función sigmoide para decidir qué valores de la memoria deben ser "leídos".
- Después, se aplica la función tangente hiperbólica (tanh) a la celda de memoria, y la salida se calcula como la combinación de ambos valores.

Desde el punto de vista de la arquitectura, las LSTM son una versión más compleja de las RNN, incorporando una celda de memoria que permite almacenar y actualizar información de manera eficiente a lo largo del tiempo [2]. Esta celda de estado, junto con las puertas de control que regulan el flujo de información, permite que las LSTM mantengan una memoria a largo plazo, superando las limitaciones de las RNN convencionales, que solo cuentan con un estado oculto para almacenar información a corto plazo [2]. En términos de rendimiento, el proceso de aprendizaje de las LSTM requiere una mayor capacidad de procesamiento debido a su arquitectura más compleja y la necesidad de optimizar múltiples parámetros [2]. Esto puede hacer que

el entrenamiento sea más costoso en términos de tiempo y recursos computacionales. Por el contrario, las RNN son más eficientes en cuanto a requerimientos computacionales, ya que su estructura más simple no exige tanto poder de procesamiento, aunque a costa de una capacidad limitada para modelar dependencias a largo plazo.

Por otro lado, existen modelos clásicos como ARIMA que son adecuados para manejar series temporales con patrones estacionales, como los datos de generación solar que suelen depender de factores como la hora del día y las estaciones del año. El modelo estacional ARIMA (SARIMA) extiende el ARIMA básico para capturar estos patrones repetitivos.

La generación solar puede presentar tendencias o variaciones estacionales significativas. ARIMA permite transformar estas series no estacionarias mediante técnicas como la diferenciación o transformaciones logarítmicas, estabilizando la media y la varianza, y facilitando su modelado. ARIMA ofrece un buen desempeño para predicciones a corto plazo, lo cual es crucial para anticipar variaciones horarias o diarias en la generación solar y optimizar el uso o almacenamiento de energía.

Existen diversos tipos de modelos de aprendizaje, como las SVR que se pueden usar para modelar las relaciones no lineales entre las variables de entrada y las de salida. Los árboles de decisión y los bosques aleatorios pueden ser útiles para predecir series temporales, ya que pueden manejar relaciones no lineales y manejar tanto variables numéricas como categóricas. Los bosques aleatorios combinan múltiples árboles para mejorar la generalización y evitar el sobreajuste.

La razón principal para utilizar redes neuronales LSTM para la predicción de la radiación solar es su capacidad para recordar información durante un largo período de tiempo con alta precisión. El estudio realizado en [4] muestra que el modelo LSTM es apropiado y competitivo para la estimación de series de tiempo de la radiación solar multivariada debido a su estructura de red. Otro estudio menciona que a diferencia de los modelos lineales, los modelos LSTM pueden capturar patrones temporales complejos y relaciones a largo plazo en los datos, que son fundamentales para las predicciones de irradiación afectadas por varios factores como las estaciones climáticas y las condiciones atmosféricas [5].

Una red neuronal recurrente (RNN) estándar tiene problemas con gradientes que desaparecen y explotan, que la red de aprendizaje profundo LSTM aborda [6]. LSTM mitiga estos problemas de gradiente mediante el uso de una arquitectura basada en memoria en su sistema de lógica de compuertas. Como resultado, genera una respuesta de gradiente aditiva al entrenamiento en lugar de una respuesta de gradiente multiplicativa. Las compuertas recurrentes, conocidas como compuertas de "olvido", se utilizan con frecuencia para permitir que LSTM almacene datos [6]. LSTM evita que los errores retropropagados desaparezcan al hacer que los errores se propaguen hacia atrás a través de múltiples capas virtuales distribuidas en el espacio [6].

V. METODOLOGÍA DE LA IMPLEMENTACIÓN

A. Selección y Obtención de Datos

1) *Explorador Solar del Ministerio de Energía*: Para este proyecto, se utilizó el Explorador Solar del Ministerio de Energía debido a la disponibilidad y confiabilidad de los datos proporcionados. La planta seleccionada fue el Parque Solar Bolero, el cual cuenta con una gran cantidad de módulos fotovoltaicos (475,000).

2) *Variables Seleccionadas*: Las variables obtenidas se encuentran documentadas en un archivo CSV (referencia: {Referencia del archivo CSV}). Las principales variables consideradas fueron:

- Radiación global: `glb` [W/m^2].
- Radiación directa: `dir` [W/m^2].
- Radiación difusa: `dif` [W/m^2].
- Radiación difusa reflejada: `sct` [W/m^2].
- Radiación difusa global horizontal: `ghi` [W/m^2].
- Radiación directa horizontal: `dirh` [W/m^2].
- Radiación difusa horizontal: `difh` [W/m^2].
- Radiación directa normal: `dni` [W/m^2].
- Temperatura: `temp` [$^{\circ}\text{C}$].
- Velocidad del viento: `vel` [m/s].
- Sombra: `shadow` ("1 o 0").
- Nubes: `cloud` ("1 o 0").

Se recopiló 113.857 datos históricos de todas estas variables climáticas y el último dato histórico es 2016-12-26 / 23:00:00.

Para justificar la selección de estas variables, se realizó un análisis de la matriz de correlación. Este análisis permite identificar relaciones entre las variables que podrían influir en la precisión del modelo predictivo. La matriz de correlación muestra los siguientes hallazgos clave:

- **Alta correlación con la variable objetivo**: La radiación global (`glb`) muestra una fuerte correlación con la generación solar (0.985) y la radiación global horizontal (`ghi`) (0.982). Esto indica que estas variables están estrechamente relacionadas y pueden proporcionar información esencial para modelar fenómenos solares y meteorológicos.
- **Relación indirecta de algunas variables**: Variables como la presencia de nubes (`cloud`) tienen una correlación baja con `glb` (0.0087), pero muestran una correlación significativa con la radiación difusa (`dif`) (0.757). Esto refleja el impacto de las nubes en la dispersión de la luz solar.
- **Efectos de la sombra y la radiación directa**: La variable sombra (`shadow`) tiene una correlación negativa fuerte con la radiación directa normal (`dni`) (-0.916), confirmando que la sombra reduce significativamente la radiación directa.
- **Relación entre sombra y nubes**: La interacción entre `shadow` y `cloud` proporciona insights importantes sobre escenarios específicos como días nublados o con topografía sombreada.
- **Temperatura y velocidad del viento**: La temperatura (`temp`) y la velocidad del viento (`vel`) tienen una

correlación positiva moderada-alta (0.71), lo que sugiere que tienden a aumentar juntas.

- **Generación solar y temperatura:** La temperatura tiene una correlación de 0.76 con la generación solar, indicando una relación positiva esperada en días más soleados y cálidos.

En la siguiente figura se muestra la matriz de correlación que resume estas relaciones.

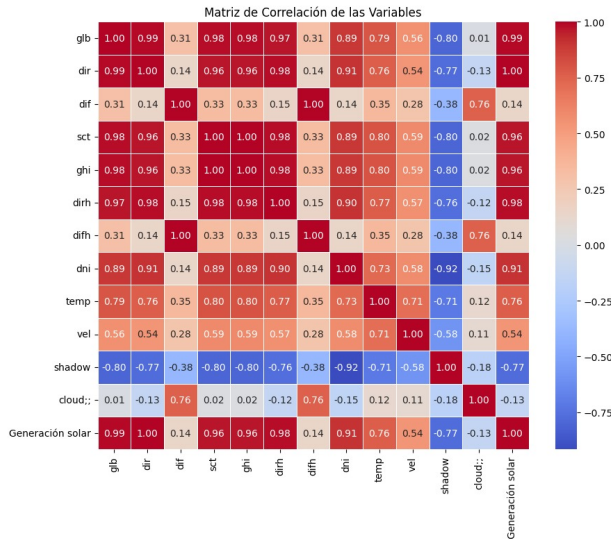


Fig. 2: Matriz de correlación de las variables seleccionadas.

El análisis confirma que las variables seleccionadas son altamente relevantes para el modelo de forecasting. Variables como ghi, dni y dif están fuertemente relacionadas con la generación solar y capturan las dinámicas del sistema. Por otro lado, shadow y cloud, aunque con correlaciones más bajas o indirectas, enriquecen el modelo al reflejar condiciones específicas que afectan la generación de energía solar.

3) *Variable de Interés:* La variable de interés para el entrenamiento inicial del modelo fue la generación solar. Para obtener esta variable, se estimó la variación de la eficiencia de los paneles fotovoltaicos y la generación esperada por estos en una hora específica. La obtención de la generación solar se realizó de la siguiente manera:

- Dimensiones de cada módulo fotovoltaico: 1.65 metros de largo por 1 metro de ancho, lo que da un área de 1.65 m² por módulo.
- Cantidad de módulos instalados: 475,000.
- Eficiencia promedio de los paneles solares: 20% (Promedio).

Si consideramos una Radiación Directa (dir) promedio de 770 [W/m²], la metodología es la siguiente:

- 1) Área total: $1.65 [m^2] \times 475,000 = 783,750 [m^2]$.
- 2) Potencia total recibida: $783,750 [m^2] \times 770 [W/m^2] = 603.49 [MW]$.

Por último, nos interesa obtener la generación solar de la planta, y para esto, utilizaremos la Radiación Directa (dir), para luego aplicar:

$$1) \text{ Generación Solar: } 603.49 [MW] \times 20\% = 120.698 [MW].$$

Es importante mencionar que los pasos descritos anteriormente corresponden para un valor particular de radiación directa, y para la implementación se generalizó esta metodología para toda la cantidad de radiaciones directas recopiladas, correspondiente a 113.857 datos históricos.

B. Procesamiento de Datos

1) *Carga de Datos:* El código para la carga de datos utiliza la biblioteca pandas de Python para procesar el archivo CSV correspondiente. A continuación, se detalla una explicación cualitativa del código:

- 1) **Importación de librerías:** El código utiliza pandas para manejar y procesar datos en formato CSV.
- 2) **Detección del inicio de los datos:** Se define una función que busca el encabezado Fecha/Hora en el archivo, lo que permite identificar correctamente la línea donde comienzan los datos relevantes.
- 3) **Carga de datos:** Utilizando `pd.read_csv`, los datos se cargan desde la línea identificada, omitiendo las filas previas.
- 4) **Limpieza de datos:**
 - Se eliminan las columnas innecesarias como Fecha/Hora.
 - Se reemplazan valores inválidos (; ;) por cadenas vacías.
 - Se convierten todas las columnas a formato numérico, gestionando errores y rellenando valores faltantes con 0.
- 5) **Cálculo de radiación en MW:**
 - Se ajustan las variables de radiación (glb, dir, dif, etc.) multiplicándolas por el área total (783,750 [m²]) y convirtiéndolas a MW.
- 6) **Aplicación de eficiencia:**
 - A la variable dir se le aplica un factor de eficiencia del 20%.
- 7) **Exportación:**
 - El DataFrame resultante se guarda en un archivo Excel (Parque_Bolero_procesado.xlsx) para futuros análisis.

Este proceso garantiza que los datos sean consistentes y estén correctamente formateados para las siguientes etapas del proyecto.

C. Preprocesamiento de Datos para el entrenamiento y validación del modelo

1) *Sección de Datos de Entrenamiento y Validación:* El código para la división de los datos utiliza una función personalizada que asegura la correcta distribución entre los conjuntos de entrenamiento, validación y prueba. A continuación, se explica cualitativamente:

- 1) **Definición de proporciones:** La función toma como parámetros las proporciones de los datos a ser asignados a cada conjunto (entrenamiento, validación y prueba),

con valores predeterminados de 80%, 10% y 10%, respectivamente.

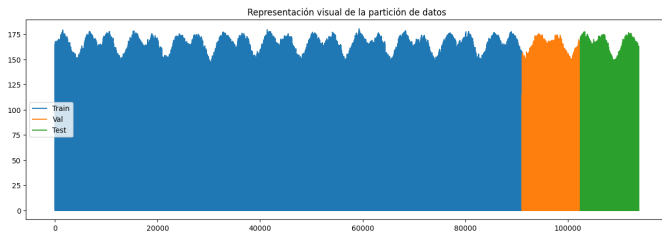


Fig. 3: Partición de datos de train, val, test

- 2) **Cálculo de tamaños:** Con base en el número total de registros, calcula el tamaño de cada conjunto.
- 3) **Partición:**

- El conjunto de entrenamiento contiene los primeros datos según el tamaño calculado.
- El conjunto de validación contiene los datos inmediatamente posteriores.
- El conjunto de prueba contiene los datos restantes.

2) *Adaptación de los Datos para Redes LSTM:* Se utiliza una función personalizada para transformar los datos en un formato adecuado para entrenar una red LSTM. La explicación cualitativa es la siguiente:

1) **Parámetros de entrada:**

- `array`: Matriz de datos, donde las filas son los registros temporales y las columnas son las variables (features).
- `input_length`: Número de pasos temporales usados como entrada.
- `output_length`: Número de pasos temporales a predecir.

2) **Generación de conjuntos X e Y:**

- Para cada registro, se toman `input_length` filas consecutivas como entrada (X).
- Se toman las `output_length` filas siguientes como salida (Y).
- Esto permite que el modelo aprenda a predecir salidas futuras basándose en datos históricos.

3) *Escalamiento de Datos:* El código para escalar los datos utiliza la biblioteca `sklearn.preprocessing.MinMaxScaler`, y su objetivo es normalizar los valores de entrada y salida al rango de -1 a 1. A continuación se detalla:

1) **Parámetros de entrada:**

- `data_input`: Diccionario que contiene los conjuntos de datos de entrada y salida para entrenamiento, validación y prueba.
- `col_ref`: Especifica la columna que contiene la variable a predecir (por ejemplo, `dir`).

2) **Inicialización:**

- Se genera un escalador (`MinMaxScaler`) por cada covariable de entrada. Esto asegura que cada variable sea escalada individualmente.

D. Configuración del Modelo LSTM

El modelo utilizado está basado en la arquitectura LSTM (Long Short-Term Memory), configurado para manejar datos temporales multivariados de entrada (12 características incluyendo la generación solar) y uni-step de salida (generación solar):

1) **Definición de la arquitectura:**

- La arquitectura incluye una capa LSTM con 128 unidades, que corresponde al tamaño del estado oculto y de las celdas de memoria.
- La entrada tiene la forma (`input_length, features`), donde `input_length` es el número de pasos temporales y `features` el número de variables de entrada.
- Una capa densa de salida con activación lineal, adecuada para problemas de regresión.

2) **Función de pérdida:**

- Se utiliza una pérdida personalizada basada en el error cuadrático medio raíz (RMSE), lo que permite evaluar el desempeño del modelo en las mismas unidades de las variables objetivo.

3) **Optimizador:**

- Se emplea el optimizador RMSprop con una tasa de aprendizaje ajustada para asegurar una convergencia eficiente del modelo.

Una vez concluida la configuración del LSTM, procede la sección de resultados de la implementación.

VI. RESULTADOS DE LA IMPLEMENTACIÓN

A. Entrenamiento y validación

Posteriormente a procesar los datos y crear la red LSTM, se continua con el entrenamiento. Simultáneamente se valida el rendimiento del modelo utilizando el indicador RMSE sobre los conjuntos de entrenamiento y validación del dataset original. El análisis del rendimiento del modelo a partir de la Figura 4 muestra un comportamiento adecuado en términos de aprendizaje y generalización. La curva azul de RMSE de entrenamiento presenta una disminución progresiva y suave a lo largo de las iteraciones, lo que indica que el modelo está optimizando correctamente la función de pérdida. Por otro lado, la curva naranja de validación también exhibe una reducción significativa en las primeras iteraciones y se estabiliza en torno a un valor constante después de aproximadamente 15 épocas.

Es importante señalar también de la figura anterior, que la RMSE de validación resulta ligeramente inferior a la de entrenamiento en la mayoría de las iteraciones. Este comportamiento indica que el modelo no se está sobreajustando a los datos de entrenamiento, lo cual es una señal positiva de generalización.

Una vez entrenada la red, volvemos a validar el desempeño del modelo con nuevos indicadores. Los resultados de la Tabla I muestran un buen desempeño del modelo LSTM en los conjuntos de entrenamiento, validación y prueba. El

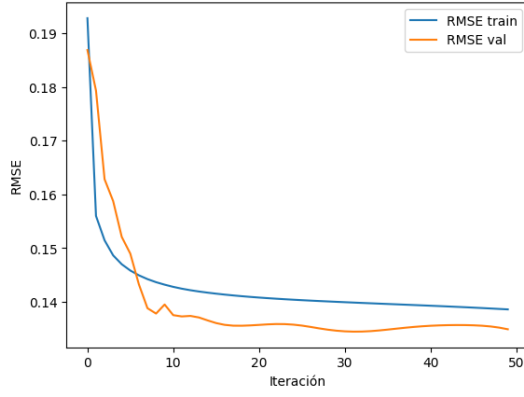


Fig. 4: Desempeño del modelo en 50 épocas

error absoluto medio (MAE) es bajo en los tres conjuntos, destacando que los valores predichos están muy cercanos a los reales. La métrica MAPE disminuye progresivamente en los conjuntos de validación y prueba, lo que sugiere que el modelo generaliza correctamente y logra predicciones más precisas en datos no vistos.

TABLE I: Resultados de desempeño del modelo LSTM

Métrica	Entrenamiento	Validación	Prueba
MAE	0.0366	0.0405	0.0391
MAPE (%)	17.25	13.41	11.32
R^2	0.9559	0.9474	0.9538
RMSE	0.07972	0.09038	0.0861

Por último, los valores de R^2 son consistentemente altos (cerca de 0.95) en todos los conjuntos, lo que indica que el modelo explica la mayor parte de la varianza de los datos. En conjunto, estos resultados confirman que el modelo tiene una buena capacidad predictiva sin mostrar signos evidentes de sobreajuste.

B. Predicciones

En la Figura 5, se observa la predicción de generación solar del modelo LSTM para las próximas 72 horas (inmediatamente posteriores al último dato real de generación obtenido), donde los valores presentan una clara periodicidad diaria. Las predicciones capturan de manera precisa los patrones cíclicos característicos de la generación solar, con picos pronunciados durante las horas de mayor radiación y valores cercanos a cero en las horas nocturnas. Este comportamiento esperado sugiere que el modelo ha logrado aprender de forma efectiva las tendencias temporales de la serie, reflejando tanto la magnitud como la frecuencia de la variabilidad diaria.

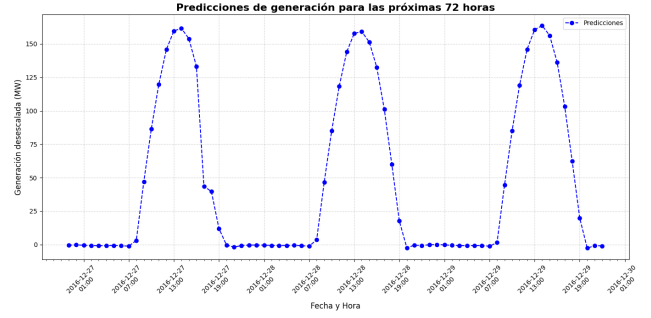


Fig. 5: Predicción de generación para las próximas 72 horas del parque solar Bolero

Por otro lado, en la Figura 6, se comparan las predicciones del modelo con los últimos 1000 puntos de datos reales de generación. La superposición de ambas curvas indica una excelente correspondencia entre las predicciones y los datos reales, donde las líneas prácticamente coinciden a lo largo de toda la serie. Los picos y valles de generación son capturados con alta precisión, lo que refleja una buena capacidad del modelo para replicar tanto la magnitud como la estructura temporal de los datos observados. Sin embargo, se notan ligeras desviaciones en algunos puntos específicos, como pequeñas diferencias en la altura de los picos o en la inclinación de los descensos, lo cual es común en modelos de predicción debido a factores externos no considerados, como variabilidad meteorológica.

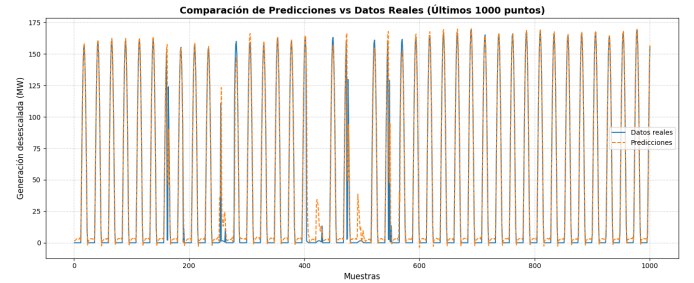


Fig. 6: Comparación entre la generación real y la pronosticada para los últimos 1000 datos del dataset

En conjunto, estos resultados sugieren que el modelo LSTM presenta un alto grado de precisión y generalización. La capacidad de predecir correctamente tanto a corto plazo (72 horas) como en comparación con los datos históricos demuestra su robustez en la captura de patrones no lineales y estacionales en la generación solar. A pesar de las leves discrepancias, el rendimiento general del modelo es satisfactorio, validando su aplicabilidad para tareas de pronóstico energético.

VII. CONCLUSIONES GENERALES

Este reporte final de proyecto ha indagado en la importancia del forecasting en la predicción de la generación de energía solar, especialmente en la penetración de ERNC en redes eléctricas modernas. La capacidad de predecir con precisión la producción solar a corto plazo es fundamental para optimizar la eficiencia operativa de las plantas fotovoltaicas y

garantizar la estabilidad del sistema eléctrico, especialmente en contextos de alta penetración de fuentes renovables.

La relevancia de implementar modelos de forecasting de generación solar radica en que, al proporcionar predicciones más precisas, los sistemas híbridos pueden gestionar de manera eficiente el almacenamiento, la distribución y la gestión de la carga de energía, reduciendo la dependencia de los combustibles fósiles y aumentando la utilización de energía renovable. La información de predicción confiable permite una mejor toma de decisiones para los operadores del sistema con respecto al almacenamiento y el despacho de energía, lo que conduce a operaciones energéticas más eficientes y sostenibles.

A través del uso de técnicas avanzadas de machine learning, como las redes neuronales LSTM (propuesta como modelo principal), este proyecto propone implementar un modelo simple de forecasting; aplicando en el proceso conocimientos aprendidos en este curso como es la programación orientada a objetos en Python.

Respecto a la implementación del modelo propuesto, la red neuronal LSTM fue entrenada utilizando datos históricos multivariados del parque solar Bolero, permitiendo capturar las tendencias temporales y no lineales inherentes a la generación solar. La estructura del modelo incluyó una capa LSTM con 128 unidades, seguida de una capa densa con activación lineal para realizar la regresión. Para garantizar la calidad del entrenamiento, se utilizó el optimizador RMSprop con una tasa de aprendizaje ajustada, y la función de pérdida fue el RMSE debido a su capacidad para medir errores en las mismas unidades que la generación solar (MW). El entrenamiento se realizó con 50 épocas y un tamaño de batch de 256, validando el modelo con un conjunto de datos separado.

Los resultados obtenidos reflejan un excelente desempeño del modelo LSTM en diferentes niveles de evaluación. En términos cuantitativos, se alcanzaron valores bajos de error absoluto medio (MAE) y error porcentual medio absoluto (MAPE), con métricas R^2 superiores al 94% tanto en entrenamiento, validación como en prueba. Esto indica que el modelo tiene una alta capacidad explicativa de la variabilidad de los datos y logra generalizar adecuadamente en conjuntos no vistos.

Las predicciones visualizadas en la gráfica de las próximas 72 horas muestran que el modelo captura con precisión los patrones cíclicos diarios, incluyendo picos pronunciados durante las horas de máxima radiación y valores cercanos a cero en la noche, lo cual es coherente con la naturaleza de la generación solar. Por otra parte, la comparación con los últimos 1000 puntos reales revela que las predicciones se ajustan de manera notable a los datos observados, con mínimas desviaciones en magnitud y tiempo, validando la robustez del modelo para predecir de manera confiable la demanda energética a lo largo del tiempo.

En conclusión, la red LSTM desarrollada demostró ser una herramienta efectiva y precisa para el pronóstico de la generación solar del parque fotovoltaico Bolero. La metodología implementada, basada en el ajuste de parámetros clave y en la validación del modelo con múltiples métricas, permitió capturar las tendencias temporales complejas y asegurar un buen rendimiento en diferentes horizontes temporales. Estos resultados consolidan el uso de redes LSTM como una solución viable para problemas de predicción energética en aplicaciones reales.

VIII. REFERENCIAS

- [1] DataScientest, "Memoria a largo plazo vs. corto plazo (LSTM)," DataScientest. [Online]. Disponible en: <https://datascientest.com/es/memoria-a-largo-plazo-a-corto-plazo-lstm/>.
- [2] Ciberseguridad Max, "Qué es LSTM (Long Short-Term Memory)," Ciberseguridad Max. [Enlace]. Disponible en: <https://ciberseguridadmax.com/lstm/>.
- [3] D. Kunalan, P. S. Krishnan and N. Permal, "Single Step Multivariate Solar Power Forecasting using Adaptive Learning Rate LSTM Model with Optimized Window Size," 2023 International Electronics Symposium (IES), Surabaya, Indonesia, 2023, pp. 70-76.
- [4] V. Gider, C. Budak, D. Izci y S. Ekinici, "Predicción diaria de la radiación solar mediante redes neuronales LSTM", Conferencia mundial sobre energía (GEC) de 2022, Batman, Turquía, 2022, págs. 168-172, doi: 10.1109/GEC55014.2022.9987055.
- [5] YI Febriansyah, DH Barus e IM Indradjaja Marcus Brunner, "Mejora de la previsión de energía solar en sistemas fotovoltaicos híbridos: análisis de series temporales basado en LSTM para la predicción de la irradiancia solar y la temperatura", Conferencia internacional de 2024 sobre tecnología y política en energía y energía eléctrica (ICTPEP), Bali, Indonesia, 2024, págs. 449-454, doi: 10.1109/ICT-PEP63827.2024.10733468.
- [6] CN Obiora, A. Ali y AN Hasan, "Encontrar el horizonte óptimo para la previsión de la irradiancia solar a corto plazo mediante una red de memoria a largo plazo (LSTM)", 11.ª Conferencia internacional sobre ciencia e ingeniería de la energía (ICPSE) de 2022, Eskisehir, Turquía, 2022, págs. 148-152, doi: 10.1109/ICPSE56329.2022.9935346.