UNIVERSIDAD DE PIURA

FACULTAD DE INGENIERÍA

Predicción de la generación eléctrica de la central solar Rubí utilizando redes neuronales LSTM

Tesis para optar el Título de
Ingeniero Mecánico - Eléctrico

José Alejandro Villanueva Mejía

Asesor(es):
Mgtr. Ing. José Hugo Fiestas Chevez

Piura, enero de 2023

NOMBRE DEL TRABAJO
Tesis_Alejandro_Villanueva.pdf

AUTOR
Alejandro Villanueva

RECUENTO DE PALABRAS
19450 Words

RECUENTO DE CARACTERES
105489 Characters

RECUENTO DE PÁGINAS
97 Pages

TAMAÑO DEL ARCHIVO
2.6MB

FECHA DE ENTREGA
Nov 19, 2022 10:00 AM GMT-5

FECHA DEL INFORME
Nov 19, 2022 10:02 AM GMT-5

15% de similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base d

- 13% Base de datos de Internet
- 2% Base de datos de publicaciones
- Base de datos de Crossref
- Base de datos de contenido publicado de Crossr
- 10% Base de datos de trabajos entregados

Excluir del Reporte de Similitud

- Material bibliográfico
- Material citado
- Material citado
- Coincidencia baja (menos de 8 palabras)

Dedicatoria

A mis padres, Homero y Rosa, y a mi hermano Eduardo.

UNIVERSITAS STUDIORUM PIURENSIS

Agradecimientos

Un especial agradecimiento a mi familia y a todos mis seres queridos por el apoyo constante en mi desarrollo profesional.

De igual forma, un agradecimiento a mi asesor Hugo Fiestas, por el apoyo en el desarrollo de esta tesis, y por resolver todas las consultas realizadas.

UNIVERSITAS STUDIORUM PIURENSIS

Resumen

El uso de energías renovables como medio de generación eléctrica ha aumentado en los últimos años, debido a sus bajos costos y por ser no contaminantes. Sin embargo, debido a la naturaleza misma de estas fuentes se tiene cierta discontinuidad de generación a lo largo del día. Para contrarrestar esta desventaja y mejorar su integración a las redes eléctricas interconectadas se han presentado distintas soluciones a partir del uso de redes neuronales con predicción en series de tiempo, los cuales permiten tener un estimado del aporte de generación eléctrica de las centrales de energía renovables.

Este estudio se enfocará en uno de estos algoritmos, llamado *Long Short Term Memory* (LSTM), el cual es usado en predicción de series de tiempo, de tal forma que se obtenga un estimado de la potencia generada a lo largo de un día de una central solar. En este caso se seleccionó la central solar Rubí, que se encuentra en el departamento de Moquegua, el cual cuenta con una instalación de 560 880 módulos fotovoltaicos de 320W, teniendo una potencia de 144,48 MW.

El contenido de la investigación aborda el marco teórico de la energía solar fotovoltaica y de las redes neuronales, el estado de los recursos energéticos renovables en el Perú y descripción de la central solar Rubí, descripción, limpieza y análisis de los datos a ingresar al modelo de predicción, y finalmente el proceso de entrenamiento y optimización de la red neuronal.

El modelo de predicción entrega resultados precisos, obteniéndose un valor de MAE de 0.0829 y de RMSE de 0.122. Adicional a estas métricas, se realizó una comparativa de la energía total producida y de las pendientes al inicio y al final de la producción tanto de los valores reales como del modelo de predicción, llegando a tener valores de 2% referente a error porcentual de diferencia en energía y valores de 12% referente a las pendientes.

De igual forma, se realizó la predicción de potencia eléctrica de siete días, ingresando los datos al set de entrenamiento de cada día posterior a su predicción, con el objetivo de ver el impacto que tiene el ingreso de datos diarios al modelo, obteniéndose como resultado que el ingreso de nuevos datos diarios no tiene un cambio en la performance del modelo.

UNIVERSITAS STUDIORUM PIURENSIS

Tabla de contenido

Introducción 17

Capítulo 1 Generalidades 19

1.1 Antecedentes 19

1.2 Objetivos 21

1.2.1 Objetivo general *21*

1.2.2 Objetivos específicos *21*

Capítulo 2 Marco teórico 23

2.1 Energía solar fotovoltaica 23

2.1.1 Radiación solar *23*

2.1.2 Paneles solares *27*

2.1.3 Sistemas fotovoltaicos conectados a la red *29*

2.1.4 Reducción de costos *31*

2.2 Redes neuronales en series de tiempo 32

2.2.1 Inteligencia artificial *33*

2.2.2 Machine Learning *33*

2.2.3 Deep Learning *37*

2.2.4 Recurrent Neural Network *40*

2.2.5 Long Short Term Memory *42*

2.2.6 Series de tiempo *43*

Capítulo 3 Desarrollo de los Recursos Energéticos Renovables (RER) en el Perú 47

3.1 Estado de los *RER en* el Perú 47

3.1.1 Potencial de generación solar en el Perú *48*

3.1.2 Subastas y planificación de RER On-Grid *49*
3.2 Central solar Rubí 55
Capítulo 4 Manejo de datos a entrenar 59
4.1 Teoría de datos 59
4.1.1 Recopilación de datos *59*
4.1.2 Limpieza de datos *60*
4.1.3 Análisis de exploración de datos *62*
4.1.4 Construcción del modelo *62*
4.2 Datos del COES 63
4.3 Datos de la NASA 65
Capítulo 5 Implementación de la red neuronal LSTM 67
5.1 Preprocesamiento de los datos a la red neuronal 68
5.2 Configuración del modelo base 76
5.3 Entrenamiento y validación de la red neuronal LSTM con el set de datos 78
5.3.1 Predicción de potencia eléctrica un día con modelo inicial *78*
5.3.2 Predicción de potencia eléctrica un día con modelo optimizado *80*
5.3.3 Predicción de potencia eléctrica de siete días *84*
5.4 Resultados 85
Conclusiones 91
Recomendaciones 93
Referencias bibliográficas 95
Apéndices 99
Apéndice A. Código LSTM 101
Anexos 107
Anexo A. Hoja informativa OSINERGMIN 109

Lista de tablas

Tabla 1. Resumen de subastas on-grid en el Perú 50
Tabla 2. Proyectos de la primera subasta 52
Tabla 3. Proyectos de la segunda subasta 53
Tabla 4. Proyectos hidroeléctricos de la tercera subasta 53
Tabla 5. Proyectos de la cuarta subasta 54
Tabla 6. Avance del desarrollo del proyecto Rubí 58
Tabla 7 Resultados del modelo inicial 79
Tabla 8. Resultados de energía y pendientes modelo inicial 80
Tabla 9 Resultados del modelo optimizado 82
Tabla 10. Resultados de energía y pendientes modelo optimizado 82
Tabla 11. Resultados del modelo optimizado con cuatro variables 84
Tabla 12. Resultados de energía y pendientes del modelo optimizado con cuatro variables 84
Tabla 13 Resultados de MAE y RMSE 86
Tabla 14. Resultados de energía y pendientes 86

UNIVERSITAS STUDIORUM PIURENSIS

Lista de figuras

Figura 1. Ángulos de posición del Sol respecto de la Tierra 24
Figura 2. Componentes de radiación solar 25
Figura 3. Ángulos de altitud, cenital y de incidencia de una superficie inclinada .. 26
Figura 4. Los tres tipos de estructuras de un material semiconductor 27
Figura 5. Curva corriente tensión y curva potencia tensión de una célula solar 29
Figura 6. Conexión de la central solar a la red interconectada nacional 31
Figura 7. Reducción del costo eléctrico en la última década con paneles solares .. 32
Figura 8. Aprendizaje supervisado de clasificación de imágenes de números 35
Figura 9. Ejemplo de algoritmo de aprendizaje supervisado de predicción 36
Figura 10. Ejemplos de clustering 37
Figura 11. Relación de AI, Machine Learning y Deep Learning 38
Figura 12. Función de activación Sigmoid 39
Figura 13. Función de activación Tanh 39
Figura 14. Función de activación ReLu 40
Figura 15. Unidad de una red neuronal recurrente a lo largo del paso de tiempo t 41
Figura 16. Partes de una celda LSTM 42
Figura 17. Serie de tiempo multivariable 43
Figura 18. Histórico de variable de temperatura 44
Figura 19. Entradas y salida de modelo LSTM 45
Figura 20. Producción de energía eléctrica por tipo de generación en el SEIN 48
Figura 21. Mapa de irradiación horizontal del Perú 49
Figura 22. Ubicación de la central solar Rubí 55

Figura 23. Paneles solares de la central fotovoltaica Rubí 56
Figura 24. Caseta de inversores y transformadores 56
Figura 25. S.E. Rubí 57
Figura 26. Diagrama unifilar 57
Figura 27. Generación eléctrica de CSF Rubí 58
Figura 28. Distribución de frecuencia de valores 61
Figura 29. Unión de dos sets de datos 61
Figura 30. Histograma para análisis de datos. 62
Figura 31. Potencia eléctrica de centrales de Enel Green Power Perú 63
Figura 32. Porcentaje de generación eléctrica por tipo de RER 64
Figura 33. Potencia eléctrica de la central solar Rubí 65
Figura 34. Toma de datos meteorológicos de la central solar Rubí 66
Figura 35. Diagrama de flujo del manejo de datos e implementación de la red neuronal 67
Figura 36. Set de datos meteorológicos 68
Figura 37. Set de datos de potencia eléctrica 68
Figura 38. Unión de sets de datos 69
Figura 39. Reemplazo de valores faltantes o no medidos a cero 70
Figura 40. Set de datos con sólo horas de generación eléctrica 70
Figura 41. Set de datos en formato CSV 71
Figura 42. Set de datos en Colab 71
Figura 43. Análisis del set de datos 72
Figura 44. Correlación de parámetros 73
Figura 45. Histórico de potencia eléctrica 73
Figura 46. Histórico de irradiancia 74
Figura 47. Histórico de índice de claridad 74
Figura 48. Histórico de ángulo cenital 75
Figura 49. Histórico de temperatura 75
Figura 50. Set de datos normalizado 76
Figura 51. Modelo de red LSTM 77

Figura 52. Reducción de error del modelo inicial 78

Figura 53. Predicción y valor real de potencia eléctrica del modelo inicial 78

Figura 54. Reducción de error del modelo optimizado 81

Figura 55. Predicción y valor real de potencia eléctrica del modelo optimizado ... 81

Figura 56. Reducción de error del modelo optimizado con cuatro variables 83

Figura 57. Predicción y valor real de potencia eléctrica del modelo optimizado con cuatro variables 83

Figura 58. Predicción y valor real de potencia eléctrica de una semana 85

UNIVERSITAS STUDIORUM PIURENSIS

Introducción

Aunque la pandemia del COVID-19 paralizó muchos proyectos energéticos durante el 2020, el uso de energías renovables continúa creciendo, representando el 80% de toda la nueva capacidad de energía eléctrica agregada en el año 2020 a nivel internacional. El apoyo en políticas y reducciones de costo podrían impulsar la participación de las energías renovables modernas en el consumo total de energía final de un 32 - 38 % para el año 2030, y las energías renovables podrían representar el 60 - 65 % de la generación de la electricidad. Sin embargo, mayores esfuerzos deben ser realizados para aumentar el uso de las energías renovables en el transporte y calefacción si se pretende cumplir los objetivos internacionales acordados (United Nations Statistics División, 2021).

Tradicionalmente, el desafío más difícil de los sistemas eléctricos es la capacidad de equilibrar tanto la generación como la demanda en todo momento. En orden de poder cumplir estas garantías, se han utilizado tecnologías convencionales sincrónicas como la generación hidroeléctrica, térmica o nuclear los cuales cumplían estos requerimientos. Por eso, la integración de paneles solares como generación eléctrica se encuentra limitada por la incertidumbre y la intermitencia de los parámetros meteorológicos involucrados en la producción, como la radiación solar y el índice de cielo despejado. Por eso, una opción es aumentar la precisión de la predicción de los parámetros mencionados (Rodríguez et al., 2021). Adicional a esto, el uso de modelos de predicción les permite a las centrales de generación eléctrica garantizar de forma anticipada la cantidad de energía que suministrarán al mercado eléctrico dentro de los horizontes de tiempo con los que trabajen.

En vista de esto, una forma de mejorar la adición de las centrales solares a las redes eléctricas peruanas es utilizando algoritmos de inteligencia artificial de predicción de generación eléctrica, los cuales utilizan los datos de las condiciones meteorológicas del lugar para predecir y tener un estimado de la energía que va a aportar al sistema interconectado.

En el primer capítulo se encuentran las generalidades del tema, incluyendo los antecedentes y los objetivos principales y específicos. Luego, en el segundo capítulo se abre paso el marco teórico, en el cual se da una descripción a la energía solar fotovoltaica y a las redes neuronales en series de tiempo.

En el tercer capítulo se presenta el desarrollo de los Recursos Energéticos Renovables (RER) en el Perú, dando un breve repaso del estado de las RER *On - Grid* en el país, las subastas realizadas y su planificación, en adición a esto se describe la central solar Rubí, a la cual se le aplicará el uso de redes neuronales para predecir su potencia eléctrica a lo largo del día.

En el cuarto capítulo se describe el manejo de datos a entrenar, desarrollando la teoría de limpieza y análisis de datos a seguir, luego se describe de los datos de la potencia eléctrica de la central solar Rubí, los cuales fueron obtenidos por parte del Comité de Operación Económica del Sistema Interconectado Nacional (COES), y por último la descripción de los datos meteorológicos obtenidos por la *National Aeronautics and Space Administration* (NASA).

En el quinto capítulo se desarrolla el entrenamiento de la red neuronal a utilizar, *Long Short Term Memory* (LSTM), iniciando con el preprocesamiento de datos a la red neuronal, la configuración del modelo y luego el entrenamiento en conjunto con el set de datos. Finalmente se obtienen los resultados de predicción de potencia eléctrica de un día, observando la performance del modelo y la ventaja de utilizar redes neuronales en el sector de energía renovable.

Capítulo 1

Generalidades

1.1 Antecedentes

La naturaleza volátil e intermitente de la energía solar en si misma presenta un desafío importante para integrarla en los sistemas de energía existentes. Debido a esto, el uso de sistemas de predicción para el pronóstico de generación eléctrica de centrales solares juega un papel importante en la solución de este problema. Con el desarrollo del *Deep Learning*, cada vez más académicos han aplicado este modelo para la predicción de series de tiempo y se obtuvo buenos resultados (Wang et al., 2019). Este tipo de tecnología ha sido utilizada en distintos rubros, de los cuales se ha podido tener excelentes predicciones, uno de estos es el sector energético, en el cual ayuda de forma directa a pronosticar con bajos márgenes de error la producción eléctrica.

En (Fouad et al., 2017) se indican los diversos factores que afectan la performance de los paneles solares y como estos factores afectan al sistema. Estos factores son el ambiente, el sistema de paneles fotovoltaicos, instalación y los factores de costo. Dentro de los factores de ambiente están la radiación solar, temperatura, acumulación de polvo y sombreado, los cuales son los que mayores efectos tienen sobre la generación eléctrica. A partir de estos datos, se puede implementar sistemas de predicción o utilizarlos para un mantenimiento correctivo.

A continuación, se mencionan algunas de las investigaciones, estudios y estados del arte referentes al tema de predicción usando redes neuronales:

(Rodríguez et al., 2021) presenta un programa de pronosticación de la radiación solar combinando diferentes ventajas del *Machine Learning* y la optimización de parámetros espaciales y temporales que permite pronosticar los siguientes 10 minutos. El paso de validación demostró que la desviación entre la radiación solar real y la pronosticada fue inferior al 4% en el 82.95% de los días examinados. En cuanto a las métricas de error, el error cuadrático medio (*root mean squared error*) fue de 50.8 W/m^2, una mejora del 11.27% en comparación con el modelo de persistencia, que fue usado como referencia. Los resultados indican que el algoritmo de predicción desarrollado

puede integrarse en los generadores fotovoltaicos para predecir su potencia de salida, promoviendo así su inclusión a la red eléctrica principal.

(Qing & Niu, 2018) proponen un novedoso esquema de predicción solar para la irradiancia solar de una hora en el futuro a partir de los datos de pronóstico del tiempo. Este estudio formula el problema de predicción como un problema estructurado de predicción de salida que predice conjuntamente múltiples salidas simultáneamente. El modelo de predicción propuesto se entrena utilizando LSTM, teniendo en cuenta la dependencia entre las horas consecutivas del mismo día. A su vez, realiza un comparativo entre el algoritmo de persistencia, la regresión lineal de mínimos cuadrados y las redes neuronales de avance de múltiples capas utilizando el algoritmo de *back propagation* para la predicción de la radiación solar. Los resultados experimentales en un conjunto de datos recopilados en la isla de Santiago, Cabo Verde, demuestran que el algoritmo propuesto supera estos algoritmos competitivos para la predicción de una única salida.

(Yin et al., 2020) este estudio presenta un método de predicción de generación de energía eólica basado en un modelo climático y en la red neuronal LSTM. Se establece un modelo de mapeo no lineal entre los elementos meteorológicos y las horas de utilización mensual de la energía eólica. Después de considerar los datos meteorológicos y la planificación de la nueva capacidad instalada, se generan los resultados de la previsión mensual de generación de energía. Un caso de estudio muestra la efectividad del método de predicción.

(Mishra et al., 2020) presentan un modelo de predicción de la energía solar a corto plazo, mediante la utilización de la capacidad de aprendizaje de la técnica de *Deep Learning* basada en LSTM y el concepto de *Wavelet Transform* (WT). En este modelo WT-LSTM, el WT se utiliza para descomponer los datos de series temporales de energía solar registrados en diferentes series de frecuencia seguido del proceso de extracción de características estadísticas. El modelo se utiliza para generar futuros valores de diferentes horizontes de tiempo (horario y diario), combinando los datos del WT junto con otros factores meteorológicos como la temperatura, velocidad del viento, presión, índice de nubes, humedad e índice de altura que sirven como valores de entrada para el algoritmo. La eficiencia del modelo es probada con distintos métodos estadísticos de performance.

(Sharadga et al., 2020) desarrollan y comparan rigurosamente varios métodos de predicción de series temporales, incluidos métodos estadísticos y los basados en inteligencia artificial, para la predicción de producción de energía fotovoltaica. Además, se investiga el efecto de la variación del horizonte temporal de predicción para todos los algoritmos. Se realizan previsiones horarias de energía solar para verificar la efectividad de diferentes modelos. Los datos utilizados en el trabajo actual comprenden 3640 horas de datos de operación tomados de una estación fotovoltaica conectada a la red de 20MW en China. La comparativa del estudio demuestra que los modelos de redes neuronales son más precisos que los modelos estadísticos sugeridos cuando son usados para predicción

de generación eléctrica en series de tiempo para paneles fotovoltaicos, además de requerir menor tiempo de computación.

(Liu et al., 2021) realizan la implementación de un algoritmo LSTM simplificado, construido sobre la arquitectura de la metodología de aprendizaje automático para pronosticar la generación de energía solar con un día de antelación. A través de los procesos de datos, ajustes de modelos, validación cruzada, evaluación de métricas y ajuste de hiper parámetros, el resultado muestra que el modelo LSTM utilizado supera al modelo de multicapa de perceptrones (MLP). La recopilación de datos se obtiene de dos zonas geográficas distintas, pero con un clima similar, dentro de los tipos de datos están los climáticos, la salida de potencia eléctrica del inversor de los paneles solares, medidores de energía y controladores.

Tal y como se observa en estas investigaciones revisadas, el uso de inteligencia artificial para la predicción de generación eléctrica da buenos resultados. Esta tecnología podría ser utilizada en distintas centrales solares o inclusive eólicas, que mediante el historial de potencia eléctrica en conjunto con las condiciones meteorológicas del lugar podría generar predicciones bastantes cercanas a las reales.

1.2 Objetivos

1.2.1 Objetivo general

Realizar la predicción de generación eléctrica de la central solar Rubí de un día utilizando el algoritmo de redes neuronales LSTM en base a los datos meteorológicos del lugar.

1.2.2 Objetivos específicos

- ✓ Seleccionar y utilizar los datos meteorológicos disponibles con una mayor relación respecto a la generación eléctrica de los paneles solares de la central.
- ✓ Entrenar y comparar los resultados obtenidos a partir de las distintas configuraciones de los hiper parámetros del modelo LSTM.

UNIVERSITAS STUDIORUM PIURENSIS

Capítulo 2

Marco teórico

2.1 Energía solar fotovoltaica

2.1.1 Radiación solar

La principal fuente de energía en nuestra tierra es la radiación solar. El valor de la radiación asciende a un valor de $1.7x\ 10^{17}$ W; 34% se reflejará, 42% será convertido en calor directamente, 23% se almacenará en vapor de agua, las ondas de agua del viento consumen 1% mientras que las plantas consumen un 0.023% (Zekry et al., 2018).

Dos ángulos definen la posición del Sol: su altitud sobre el horizonte; y su rumbo en el horizonte, que se llama el ángulo de acimut. Un tercer ángulo, el ángulo cenital, se utiliza a menudo en lugar de la altitud. Mientras que la altitud es la distancia angular sobre el horizonte, el ángulo cenital es la distancia angular desde una línea perpendicular a la superficie. La altitud es cero en el amanecer y el atardecer, mientras que su mayor valor es en el mediodía solar. En el mediodía solar, el Sol se encuentra al sur en el horizonte en el hemisferio norte y al norte en el hemisferio sur. Por la mañana, el Sol está al sureste, y en la tarde está en el suroeste (Bonan, 2015).

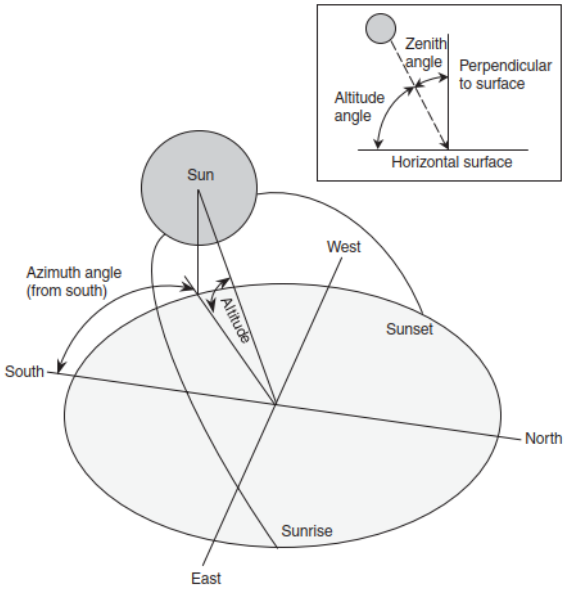
El ángulo cenital y el ángulo de altitud varían según la latitud, la época del año y la hora del día. Esto se ve expresado en la siguiente ecuación:

$$\cos Z = \sin B = \sin \emptyset \sin \delta + \cos \emptyset \cos \delta \cos h \qquad (1)$$

Donde Z es el ángulo cenital, B es el ángulo de altitud, $\emptyset$ es la latitud, δ es la declinación solar, y h es el ángulo solar horario. El ángulo de declinación solar varía con el día del año, este se forma mediante una línea imaginaria que une al Sol y la tierra. El ángulo solar horario mide el tiempo del día, debido a que la tierra gira en su eje de 360° en 24 horas, una hora equivale a 15° de longitud, y el ángulo horario varía en 15° por cada hora antes o después del mediodía solar (Bonan, 2015).

En la figura 1 se tiene una representación gráfica de los ángulos de acimut, altitud y cenital:

Figura 1. Ángulos de posición del Sol respecto de la Tierra



Nota. **Adaptado de Bonan (2015).**

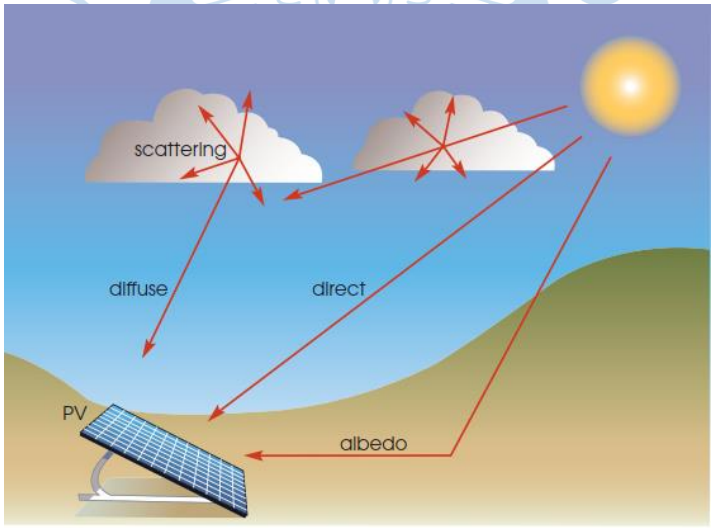
La radiación emitida por el Sol atraviesa el espacio vacío en todas direcciones. No sufre pérdidas apreciables por la interacción con otros medios materiales en el camino. Sin embargo, la irradiancia solar, definida como la densidad de flujo radiante solar, es atenuada de acuerdo con el cuadrado de la distancia. Parte de esta irradiancia solar es interceptada por el planeta Tierra, la cual varía a lo largo del año debido a la excentricidad de la elipse que recorre el planeta (Perpiñán, 2013). Esto indica que los valores de irradiación se ven afectados por el nivel de acercamiento que se tiene respecto del Sol, además de la zona del planeta, en la cual varía el ángulo con el cual llega la energía.

La unidad de medida de la irradiación es el Watt por metro cuadrado (W/m^2). Cuando la irradiancia (una cantidad instantánea) se acumula con el tiempo, se convierte en irradiación (a veces también denominada "exposición radiante"). En principio, debe expresarse en la unidad SI adecuada; es decir, kJ/m^2 o MJ/m^2. Sin embargo, en la mayoría de las aplicaciones de energía solar, la unidad diaria de kWh/m^2 se usa con más frecuencia, por conveniencia ($1\ kWh = 3.6MJ$). En muchos casos, las irradiaciones horarias o subhorarias se notifican en términos de irradiancia media, por lo que se expresan en W/m^2 (Martín & Dominguez, 2019).

Los paneles solares son instalados por lo general al nivel o cerca del nivel del suelo, por lo general reciben radiación solar directa e indirecta. La radiación directa es la que va dirigida directamente al panel solar, sin encontrarse con otro objetivo durante su recorrido. Mientras que la radiación indirecta se ve compuesta por la radiación difusa, el cual está conformado por la luz dispersada por las nubes y partículas de polvo de la atmósfera, y el albedo, que representa la luz reflejada por la tierra u objetos como los árboles o edificios (Lynn, 2010).

La energía eléctrica que se obtiene de un panel solar depende del efecto combinado de la irradiación directa, difusa y albedo. Dependiendo del clima, uno de estos componentes se vuelve más predominante que el otro; por ejemplo, si es un día nublado, el componente difuso será el que mayor porcentaje de participación tenga (siendo que las nubes bloquean o atenúan el pase directo de la radiación). El componente del albedo por lo general tiene un aporte pequeño; sin embargo, se vuelve significante en lugares como los Alpes suizos debido a fuertes reflejos de la nieve (Lynn, 2010). Se puede observar en la figura 2 los componentes de la radiación y como llegan a un panel solar.

Figura 2. Componentes de radiación solar



Nota. **Adaptado de Lynn (2010).**

La composición de la irradiación que llega al planeta se encuentra expresada de la siguiente manera:

$$GHI = DHI + DIF = DNI \cos Z + DIF \quad (2)$$

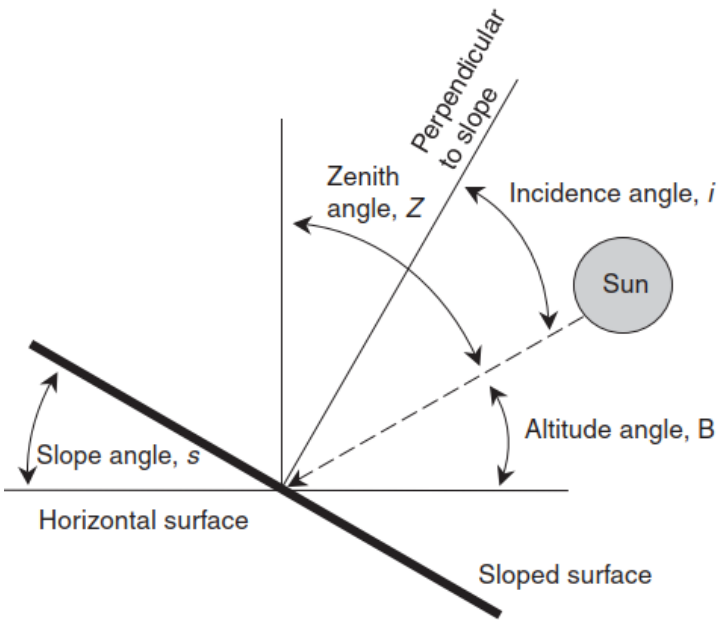
La irradiancia horizontal global (Global horizontal irradiance o por sus siglas en inglés GHI), es el resultado de la suma de sus dos componentes, la irradiancia horizontal directa (Direct horizontal irradiance o por sus siglas en inglés DHI), más la irradiancia horizontal difusa (Diffuse horizontal irradiance o por sus siglas en inglés DIF) (Martín & Dominguez, 2019).

A su vez, un factor que influye en la irradiancia sobre una superficie o cuerpo es su grado de inclinación. Según lo revisado en (Bonan, 2015), la radiación directa del haz sobre una superficie inclinada depende del ángulo cenital del Sol, el ángulo de inclinación, la dirección del Sol y la dirección de la superficie inclinada. Una superficie recibe la máxima intensidad cuando se orienta perpendicularmente a los rayos del sol. En una superficie inclinada, la desviación angular de la perpendicular debe ajustarse a la inclinación de la superficie. El ángulo de incidencia es el ángulo entre el rayo del Sol y una línea imaginaria perpendicular a la pendiente. Este ángulo es:

$$\cos i = \cos s \cos Z + \sin s \sin Z \, (A_{sol} - A_{pendiente}) \qquad (3)$$

Donde s es el ángulo de pendiente y $A_{pendiente}$ es el acimut de la pendiente. Esta relación de los ángulos mencionados se puede apreciar de mejor forma en la figura 3:

Figura 3. Ángulos de altitud, cenital y de incidencia de una superficie inclinada



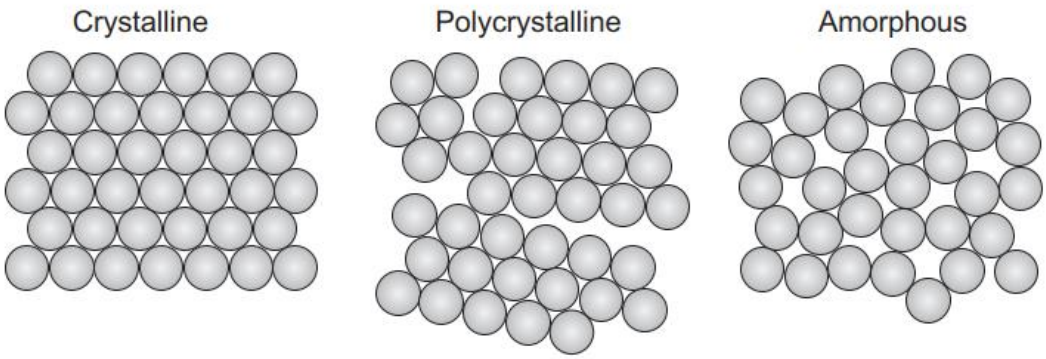
Nota. **Adaptado de Bonan (2015).**

2.1.2 Paneles solares

Los paneles solares son un conjunto de células fotovoltaicas, estas células son semiconductores, una clase de material que tiene conductividad eléctrica entre el metal y los aisladores. En consecuencia, sus características eléctricas pueden controlarse por dopaje y por luz. Se pueden clasificar de la siguiente manera: a partir de su estructura química en forma elemental de Silicio (Si) y Germanio (Ge), o en forma molecular como los semiconductores orgánicos o los perovskitas que son moléculas orgánicas híbridas metálicas. Todos los semiconductores tienen en común propiedades distinguidas que gobiernan su performance. Las principales propiedades que controlan su comportamiento eléctrico, se basan en sus arreglos atómicos (Zekry et al., 2018).

Existen tres tipos de arreglos atómicos que forman las celdas solares de silicio cristalino actualmente: monocristalino, policristalino y amorfo. La figura 4 representa sus diferentes estructuras.

Figura 4. Los tres tipos de estructuras de un material semiconductor



Nota. **Adaptado de Zekry et al. (2018).**

El primero es el monocristalino; el cual es el más eficaz de todos, está hecho de una capa delgada rebanada u oblea de un monocristal grande obtenido de un silicio fundido puro. Las obleas circulares, a menudo de cinco o seis pulgadas (15 cm) de diámetro, tienen una apariencia plateada suave y normalmente se recortan en una forma pseudo cuadrada o hexagonal para que se puedan colocar más en un módulo. Se usan barras colectoras para conducir la corriente eléctrica lejos de las celdas que tienen una estructura cristalina altamente ordenada con propiedades uniformes y predecibles. Sin embargo, requieren procesos de fabricación cuidadosos y costosos, incluido el "dopaje" con pequeñas cantidades de otros elementos para producir las características eléctricas requeridas

(Lynn, 2010). La eficiencia comerciales de este tipo de panel solar es aproximadamente del 21.2% (Green et al., 2020).

Luego está el policristalino, este tipo de célula también es producido a partir del silicio fundido puro, pero utilizando otro proceso de fundición. A medida que el silicio se enfría, se solidifica como un gran multicristal irregular que luego es cortado en rebanadas finas cuadradas o rectangulares para hacer individuales células. Su estructura cristalina, al ser aleatoria, es menos ideal que la del material monocristalino y proporciona eficiencias de celda ligeramente inferiores. Sin embargo, esta desventaja se compensa debido a que representa un costo más bajo. Las células y módulos de este tipo a menudo se ven claramente azules, con una apariencia brillante y escamosa (Lynn, 2010). El valor de la eficiencia de los paneles policristalinos ha mejorado en los últimos años, llegan a tener valores de 18.4% (Green et al., 2020).

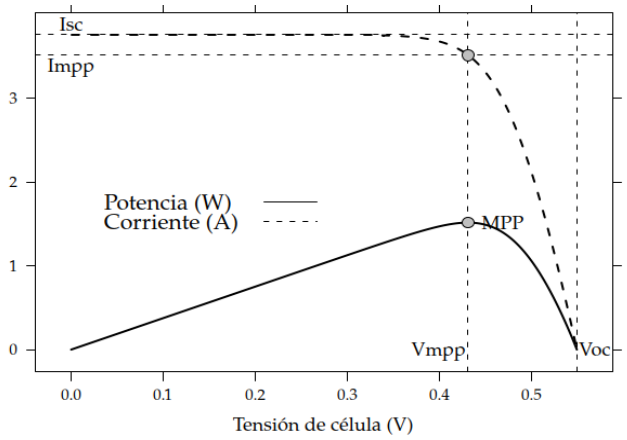
El tercer tipo es el amorfo. La mayoría de las personas han encontrado pequeñas celdas de silicio amorfo (a-Si) en productos de consumo que funcionan con energía solar, como relojes y calculadoras, que se introdujeron por primera vez en la década de 1980 (Lynn, 2010). Las células amorfas son más baratas que las otras dos, pero tienen eficiencias más bajas, aproximadamente de un 10% (Green et al., 2020).

Debido al principio de funcionamiento de una célula solar, los paneles solares funcionan en corriente continúa. Estos equipos representan su generación de potencia mediante el uso de una curva, llamada curva de intensidad – voltaje o también I-V. Todos los fabricantes de paneles solares utilizan esta curva para representar los valores de voltaje, corriente y potencia que cuentan sus productos. El aumento de temperatura afecta el rendimiento de los paneles, por lo que es común mostrar distintas curvas de un mismo modelo de panel solar con la diferencia en la temperatura de funcionamiento, siendo inversamente proporcional a la temperatura.

En la curva se encuentran los valores máximos de corriente (corriente de cortocircuito) y voltaje (voltaje de circuito abierto). La corriente de cortocircuito es el valor que asume la corriente cuando la tensión es igual a cero en las celdas solares, se representa como I_{sc}. El voltaje de circuito abierto, el cual se representa como la tensión máxima, se tiene cuando el amperaje es igual a cero, el rango en el que se encuentra va de 0.5 V hasta 0.6 V por cada célula, esta se representa como V_{oc} (Lynn, 2010). La potencia máxima es dada y graficada dentro de la curva I-V por el fabricante, la cual se obtiene a partir de la corriente y voltaje de máxima potencia.

Un ejemplo de esta curva se puede apreciar en la figura 5, donde se considera una temperatura ambiente de 20°C e Irradiancia de 800 W/m^2.

Figura 5. Curva corriente tensión y curva potencia tensión de una célula solar



Nota. **Adaptado de Perpiñán (2013).**

2.1.3 Sistemas fotovoltaicos conectados a la red

Un sistema fotovoltaico es el conjunto de equipos eléctricos y electrónicos que producen energía eléctrica a partir de la radiación solar. El principal componente de este sistema es el módulo fotovoltaico, a su vez compuesto por células capaces de transformar la energía luminosa incidente en energía eléctrica de corriente continua. El resto de los equipos incluidos en un sistema fotovoltaico depende en gran medida de la aplicación a la que está destinado. A grandes rasgos los sistemas fotovoltaicos pueden clasificarse en tres grandes grupos: conectados a la red, autónomos y de bombeo (Perpiñán, 2013). Para este trabajo, se enfocará en los sistemas conectados a la red.

Los sistemas fotovoltaicos conectados a la red son sistemas cuya función es producir energía eléctrica en condiciones adecuadas para poder ser inyectada en la red convencional. Para una central solar, toda la energía generada va directamente a integrarse a la red nacional, por lo que no hay una repartición a realizarse de la energía. En este caso, el sistema debe diseñarse para la entrega de energía a la red, asegurando un correcto nivel de confiabilidad y eficiencia.

El diseño de estos sistemas se puede realizar con diferentes complejidades. El rendimiento de los componentes del sistema fotovoltaico puede cambiar debido a las condiciones ambientales y los cambios climáticos. Estos cambios en el rendimiento del sistema pueden alterar significativamente el diseño principal. Por lo tanto, estas consideraciones deben tomarse en cuenta para el desarrollo del diseño de una central fotovoltaica, de igual forma, se debe seguir con los estándares nacionales de cada país (Aghaei et al., 2020).

La generación eléctrica por parte de las centrales solares se da en corriente continua (CC), por lo que es requerido transformarlo a alterna (AC) para su integración a la red. Para esto, se utiliza un inversor, el cual tiene como alimentación CC y una salida AC

trifásica a 415 V o un valor superior (Satpathy & Pamuru, 2021). Sin embargo, esto dependerá del país y la central solar, en el caso analizado en esta investigación, la central solar Rubí cuenta con inversores de 0.4 kV AC.

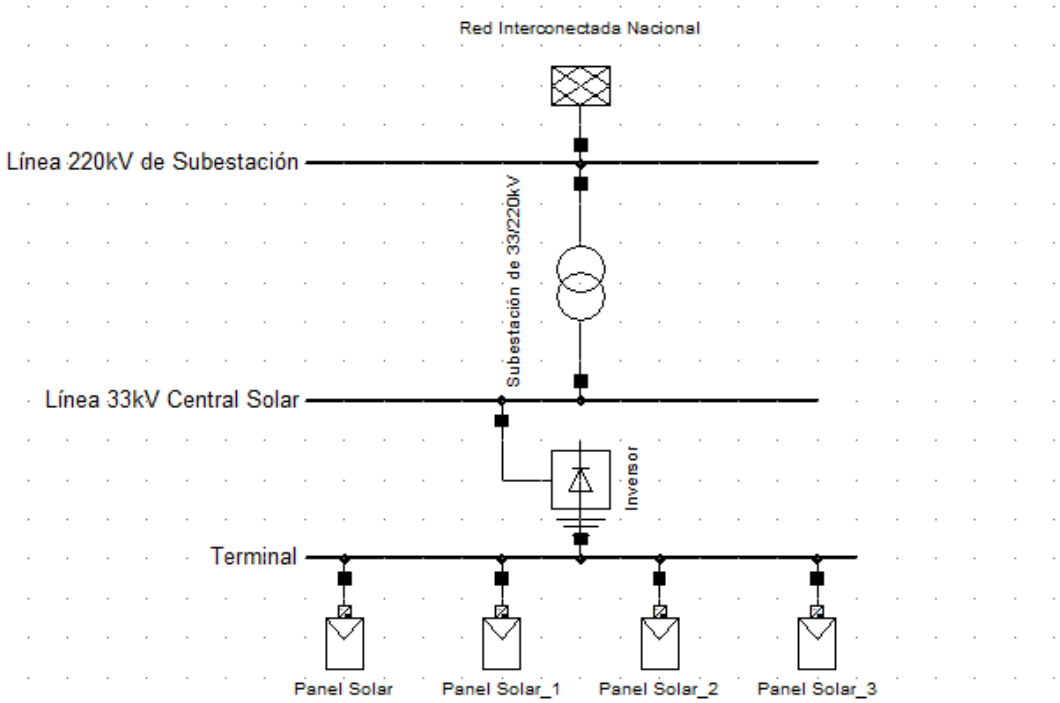
El inversor, como se menciona, es un elemento clave para realizar la conversión de la corriente. Estos equipos utilizan electrónica avanzada para producir la corriente alterna en la frecuencia y voltaje correctos para que coincida con el suministro de red. Si bien se puede utilizar una unidad cuando se trata de un uso doméstico, las unidades requeridas van en aumento cuando es mayor la potencia, eficiencia, confiabilidad y seguridad, las cuales son las principales preocupaciones del diseñador de sistemas. Se requieren grandes cantidades dentro de una central solar, por lo que estos dispositivos deben funcionar en el intervalo de potencia que se espera de un equipo fotovoltaico, por lo general, se usa un seguimiento de potencias máxima (*maximum power point tracking* o por sus siglas en inglés MPPT) para optimizar el rendimiento de energía de salida (Lynn, 2010).

En el caso de las centrales solares, su aporte de energía es alto, por lo que luego de haber convertido a alterna la corriente con los inversores, se conecta a un transformador, por lo general estos se encuentran dentro de una subestación, el cual permite elevar el voltaje e integrarlo al nivel de la línea de transmisión.

Un transformador es un dispositivo estático que transforma la energía eléctrica en alterna de un voltaje a otro, mediante el principio de inducción electromagnética, manteniendo la misma frecuencia en la que se encuentra. Existen distintos tipos de transformador, en el caso de los sistemas fotovoltaicos se utiliza los transformadores elevadores o de potencia. Los transformadores de potencia aumentan los voltajes de 22/11/33 a 66/110/132/220 kV para un sistema interconectado. Se puede utilizar transformadores de devanados múltiples, ya que cuentan con una ventaja de poder conectar con más inversores al mismo tiempo para utilizar menos espacio (Satpathy & Pamuru, 2021).

A partir del transformador, la energía pasa a conectarse a la red interconectada nacional. El paso de la energía eléctrica de los paneles hacia el transformador se puede apreciar en la figura 6:

Figura 6. Conexión de la central solar a la red interconectada nacional

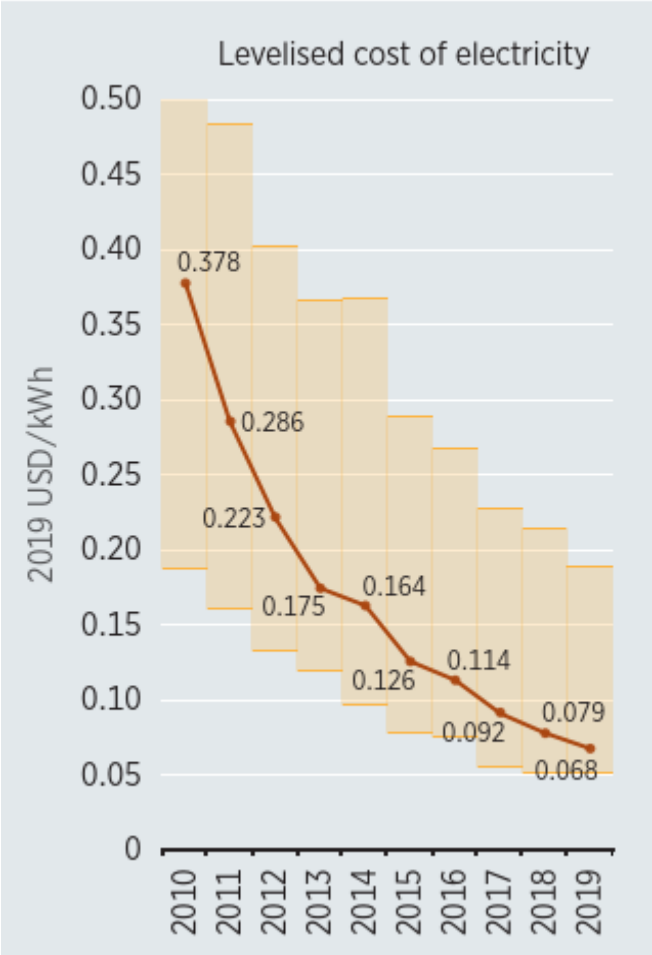


2.1.4 Reducción de costos

La rápida caída en los costos totales instalados, el incremento de factores de capacidad y la caída de los costos de operación y mantenimiento, han contribuido a la notable reducción del costo de electricidad por parte de la energía solar fotovoltaica y su competitividad económica. Para realizar una comparativa entre las distintas fuentes de energía, se usará el término coste nivelado de energía (*levelized cost of energy* o por sus siglas en inglés LCOE), el cual representa el valor del coste total de construir y operar una instalación generadora a lo largo de toda su vida útil. El LCOE de las plantas fotovoltaicas a gran escala disminuyó un 82% entre los años 2010 y 2019, de alrededor de 0.378 USD/kWh a solo 0.068 USD/kWh. Esta estimación de 2019 representa una disminución interanual del 13% con respecto al 2018 (IRENA, 2020).

Esta reducción demuestra que las fuentes de energía alternativas son más beneficiosas económicamente, por lo que las empresas generadoras se verán más inclinadas a apostar por este tipo de soluciones que las tradicionales. La reducción del costo eléctrico se observa en la figura 7.

Figura 7. Reducción del costo eléctrico en la última década con paneles solares



Nota. **Adaptado de IRENA (2020).**

2.2 Redes neuronales en series de tiempo

La cantidad de datos que en la actualidad se tienen en el mundo va en aumento exponencial, y las distintas aplicaciones que se les puede dar, de igual forma. La herramienta más popular es el uso de modelos de inteligencia artificial. Esta tecnología puede ser aplicada para el reconocimiento de imágenes, generación de texto, predicción, detección de spam o malware y una larga lista de usos. Todas estas aplicaciones son utilizadas en distintos campos como el sector eléctrico, medicina, manufactura, *marketing*, agricultura, finanzas, entre otras.

En consecuencia, a la larga lista de conceptos y terminología de esa área, es conveniente aclarar sus definiciones, una breve introducción de la historia de la inteligencia artificial y a que campo o subcampo pertenecen. Por lo que se procederá a definir los conceptos claves y consecuentemente se centrará en el modelo a utilizar para las predicciones de series de tiempo.

2.2.1 Inteligencia artificial

Los sistemas de inteligencia artificial o AI (*artificial intelligence*) por sus siglas en inglés, son unos sistemas de *software* (y posiblemente también de hardware) diseñados por humanos que, dado un objetivo complejo, actúan en la dimensión física o digital al percibir su entorno a través de la adquisición de datos, interpretando los datos recopilados estructurados o no estructurados, razonando el conocimiento, procesando la información, derivando de estos datos y decidiendo las mejores opciones a tomar para lograr el objetivo dado. Los sistemas de AI pueden usar reglas simbólicas o aprender un modelo numérico, y también pueden adaptar su comportamiento analizando cómo el entorno se ve afectado por sus acciones anteriores (European Comimission, 2019).

Los inicios de la inteligencia artificial datan del siglo pasado. El primer trabajo que se reconoce como AI fue realizado por Warren McCulloch y Walter Pitts. Propusieron un modelo de neuronas artificiales en el que cada neurona se caracteriza por estar "encendida" o "apagada", con un cambio de "encendido" en respuesta a la estimulación de un número suficiente de neuronas vecinas. Además, demostraron que cualquier función computable podía ser computada por alguna red de neuronas conectadas, y que todos los conectores lógicos podrían implementarse mediante estructuras de red simples. De igual forma, dos estudiantes de pregrado en Harvard, Marvin Minsky y Dean Edmonds, construyeron la primera computadora de red neuronal en 1950. El SNARC, como se le llamó, usó 3000 tubos de vacío y un mecanismo de piloto automático sobrante de un bombardero B-24 para simular una red de 40 neuronas (Stuart & Peter, 2021).

Durante bastante tiempo, muchos expertos creyeron que la inteligencia artificial a nivel humano podría lograrse haciendo que los programadores crearan a mano un conjunto suficientemente grande de reglas explícitas para manipular conocimiento. Este enfoque se conoce como AI simbólica y fue el paradigma dominante en la IA desde la década de 1950 hasta finales de la década de 1980. Alcanzó su máxima popularidad durante el auge de los sistemas expertos de la década de 1980 (Chollet, 2018).

Aunque la AI simbólica demostró ser adecuada para resolver problemas lógicos bien definidos, como jugar ajedrez, resultó ser poco efectivo para descubrir reglas implícitas para resolver problemas más complejos y confusos, como la clasificación de imágenes, el reconocimiento de voz y la traducción de idiomas. Un nuevo enfoque se abrió paso que terminó por desplazarla, el Machine Learning (Chollet, 2018).

2.2.2 Machine Learning

El *Machine Learning* o aprendizaje automático es un marco algorítmico unificado diseñado para identificar modelos computacionales que describen con precisión los datos empíricos y los fenómenos subyacentes, con poca o ninguna participación humana. Si bien todavía es una disciplina joven con mucho más descubrimiento en espera de lo que se conoce actualmente, hoy en día el *Machine Learning* se puede usar para enseñar a las computadoras a realizar una amplia gama de tareas útiles, incluida la detección

automática de objetos en imágenes, reconocimiento de voz, descubrimiento de conocimientos en las ciencias médicas y análisis predictivo, por nombrar algunos (Watt et al., 2020).

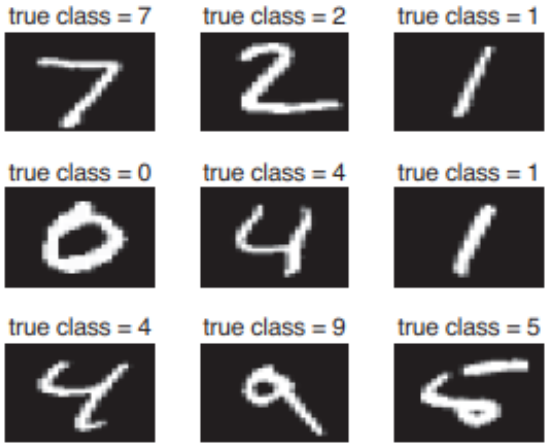
Uno podría plantearse la pregunta ¿Por qué quisiéramos que una máquina aprenda? ¿Por qué no solo programarlo para realizar específicamente una actividad? En (Stuart & Peter, 2021), dan dos razones principales. Primero, los programadores no pueden anticipar todas las situaciones futuras posibles. Por ejemplo, un robot diseñado para navegar laberintos debe aprender el diseño de cada nuevo laberinto que encuentra; un programa para predecir los precios del mercado de valores debe aprender a adaptarse cuando las condiciones cambian de auge a caída. En segundo lugar, a veces los programadores no tienen idea de cómo programar una solución por su propia cuenta, por lo que el uso de algoritmos de *Machine Learning* facilitan la creación de estos modelos.

Los algoritmos de *Machine Learning* pueden ser clasificados en 3 categorías: aprendizaje supervisado (*supervised learning*), aprendizaje no supervisado (*unsupervised learning*) y aprendizaje por refuerzo (*reinforcement learning*).

2.2.2.1 Aprendizaje supervisado. Los problemas de aprendizaje supervisado se refieren al aprendizaje automático de reglas computacionales que involucran relaciones de entrada y salida. Aplicable a una amplia gama de situaciones y tipos de datos, este tipo de problema se presenta en dos formas, denominadas clasificación y regresión, según la forma numérica general de la salida (Watt et al., 2020).

Para la clasificación, el objetivo es aprender un mapeo de las entradas x a las salidas, habiendo clasificación binaria cuando el número de clases es de 2, cuando el número de clases es mayor a dos se denomina clasificación multiclase. Si las etiquetas de clase no son mutuamente excluyentes (por ejemplo, alguien puede clasificarse como alto y fuerte), se describe como clasificación de múltiples etiquetas, pero esto se ve mejor como una predicción de múltiples etiquetas de clase binaria relacionadas (el llamado modelo de salida múltiple). Cuando se usa el término clasificación, no se hace referencia a una clasificación multiclase con una única salida, a menos que se indique lo contrario (Murphy, 2012). Un ejemplo claro de clasificación es el reconocimiento de objetos en imágenes, diferenciando distintos objetos, animales o personas a partir de los rasgos que detecta el algoritmo en su entrenamiento, un ejemplo de clasificación se encuentra en la figura 8.

Figura 8. Aprendizaje supervisado de clasificación de imágenes de números



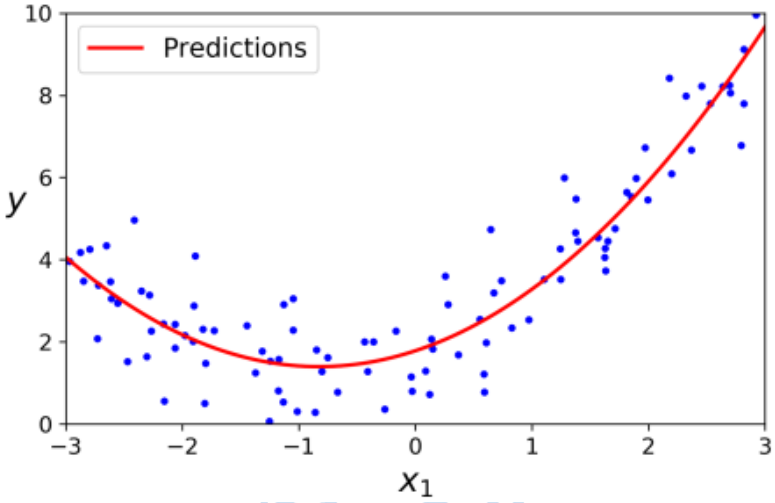
Nota. **Adaptado de Murphy (2012).**

En el caso de la regresión, se trata de un modelo que, a partir del entrenamiento de un set de datos, se genere relaciones entre estos valores y reconozca patrones, dando como resultado la predicción de los valores de salida.

Pueden surgir varias extensiones de este problema, como tener entradas de alta dimensión, valores atípicos, respuestas no uniformes, etc. Se podría decir que la diferencia con la clasificación es que la variable de salida es continua (Murphy, 2012). Este tipo de modelo puede ser aplicado al valor de la temperatura de una localidad a partir del histórico de sus distintos datos meteorológicos.

Un ejemplo de predicción se puede apreciar en la figura 9, mostrando el valor de predicción del algoritmo frente al valor real.

Figura 9. Ejemplo de algoritmo de aprendizaje supervisado de predicción



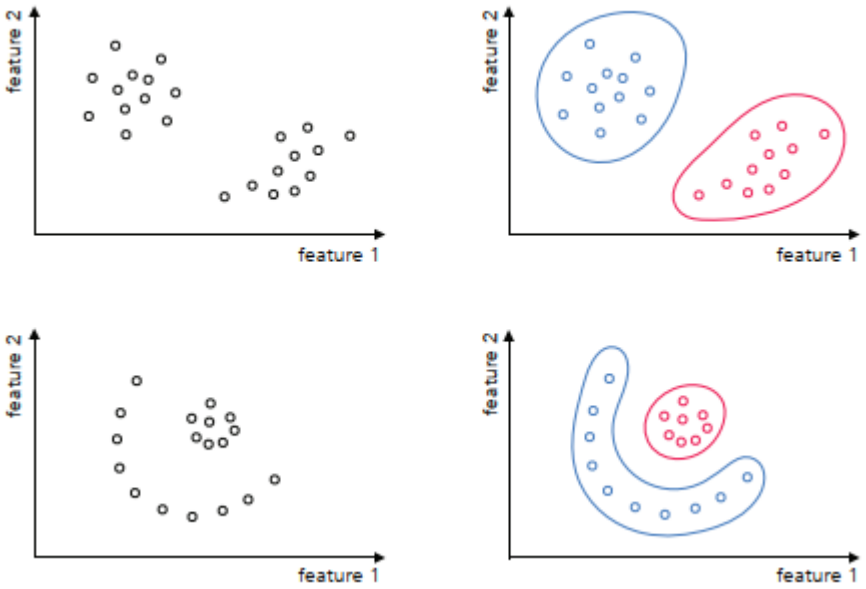
Nota. **Adaptado de Géron (2011).**

2.2.2.2 Aprendizaje no supervisado. Se ocupa del aprendizaje automático de las reglas computacionales que describen únicamente los datos de entrada. A menudo, estas reglas se aprenden para simplificar un conjunto de datos para permitir aprendizaje supervisado más fácil o para el análisis y la interpretación del analista. Dos problemas fundamentales de esta categoría son, la reducción de dimensiones y el agrupamiento, permiten la simplificación de un conjunto de datos a través de dos caminos naturales: ya sea reduciendo la dimensión ambiental de los datos de entrada o determinado un pequeño número de representantes que describan adecuadamente la diversidad de un conjunto más grande de datos (Watt et al., 2020).

Un ejemplo claro de aprendizaje no supervisado es el de agrupación o llamado clustering. El clustering tiene como objetivo identificar la estructura subyacente bruta en un conjunto de datos de entrada al agrupar puntos que comparten alguna característica estructural, por ejemplo, la proximidad entre sí en el espacio de características, lo que ayuda a organizar o resumir mejor un conjunto de datos de entrenamiento para el análisis por parte del usuario. La estructura de los datos puede variar enormemente, con datos que caen en cúmulos globulares o a lo largo de variedades no lineales (Watt et al., 2020).

Este tipo de estructura de agrupación de datos que se da en el clustering se ve reflejada en la figura 10, teniendo estas distintas formas de agrupaciones en base a sus características similares.

Figura 10. Ejemplos de clustering



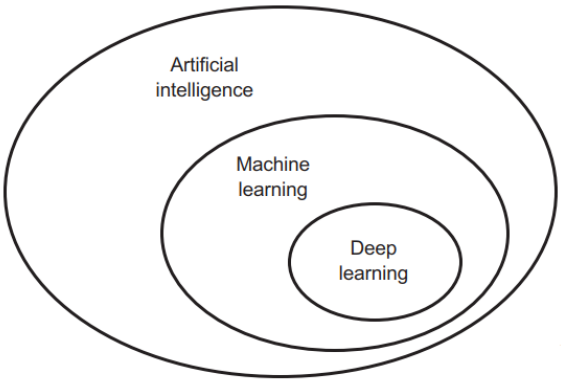
Nota. **Adaptado de Watt et al. (2020) .**

2.2.2.3 Aprendizaje por refuerzo. En el aprendizaje por refuerzo o llamado *reinforcement learning*, se utiliza un algoritmo, el cual funciona como un agente que debe tomar una serie de decisiones que proporcionan recompensas y a partir del resultado, revisar la calidad de su comportamiento. El objetivo es optimizar la suma de futuras recompensas. Todas estas funciones se pueden representar mediante gráficos de cálculo (Stuart & Peter, 2021). Por ejemplo, este modelo de *Machine Learning* se popularizó después que fuese utilizado por el equipo de Google *DeepMind* y fuese capaz de derrotar al campeón mundial de Go, un juego de estrategias.

2.2.3 Deep Learning

El aprendizaje profundo o *Deep Learning* es un subcampo específico del *Machine Learning*, una forma de aprender representaciones o características a partir de datos que pone énfasis en aprender capas sucesivas de representaciones. Lo profundo en el *Deep Learning* no es una referencia a ningún tipo de comprensión más profunda lograda por el enfoque; más bien, representa la idea de representación de sucesivas capas. Modelos modernos de este subcampo implican decenas o cientos de sucesivas capas de aprendizaje y todas estas aprenden de manera automática de la exposición de los datos de entrenamiento (Chollet, 2018). La figura 11 representa la relación de los términos AI, *Machine Learning* y *Deep Learning*.

Figura 11. Relación de AI, Machine Learning y Deep Learning



Nota. **Adaptado de Chollet (2018).**

El *Deep Learning* tiene su origen en los primeros trabajos que intentaron modelar las redes de neuronas del cerebro con circuitos computacionales. Por esta razón, las redes entrenadas por métodos de *Deep Learning* son comúnmente llamadas redes neuronales o *neural networks*, aunque el parecido con las células y las redes neuronales es superficial (Stuart & Peter, 2021).

Existen distintos tipos de *Deep Learning*, dos de estos son: la primera es una red *feedforward*, el cual tiene conexiones en solo una dirección, de tal forma que la información fluye a través de la red desde los nodos de entrada a los nodos de salida y no hay bucles. La segunda red es la recurrent network, la cual retroalimenta sus salidas intermedias o finales en sus propias entradas. Esto significa que los valores de la señal dentro de la red forman un sistema dinámico que tiene estado interno o memoria (Stuart & Peter, 2021).

Cada nodo dentro de una red es llamada unidad. Una unidad calcula la suma ponderada de las entradas de los nodos predecesores y luego aplica funciones no lineales que producen una salida. Estas funciones son llamadas funciones de activación (Stuart & Peter, 2021). Estas funciones de activación son las que permiten darles un peso a las características detectadas por el algoritmo, de tal forma que pueda categorizar o identificar la relación que hay del set de datos.

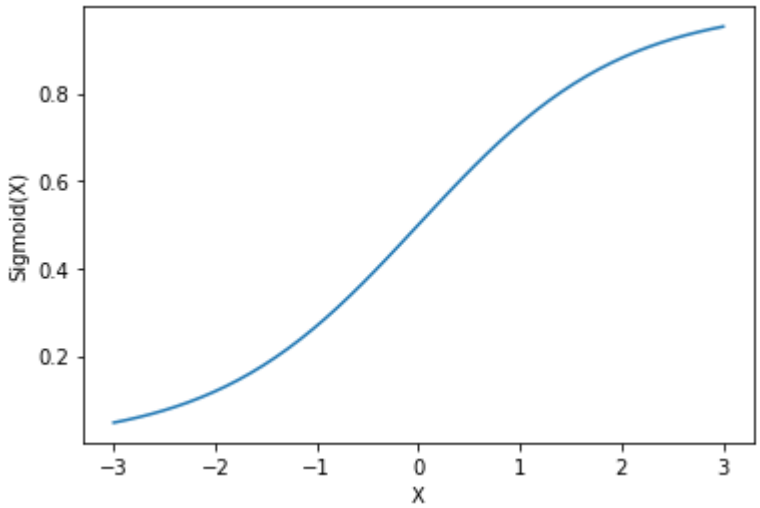
Existen distintos tipos de funciones de activación que son utilizadas por las unidades de estos algoritmos, a continuación se mencionará unas de las importantes.

La función de activación Sigmoid se define con la siguiente fórmula:

$$S(x) = \frac{1}{1+e^{-x}} \tag{4}$$

La función Sigmoid se ve representada en la figura 12.

Figura 12. Función de activación Sigmoid

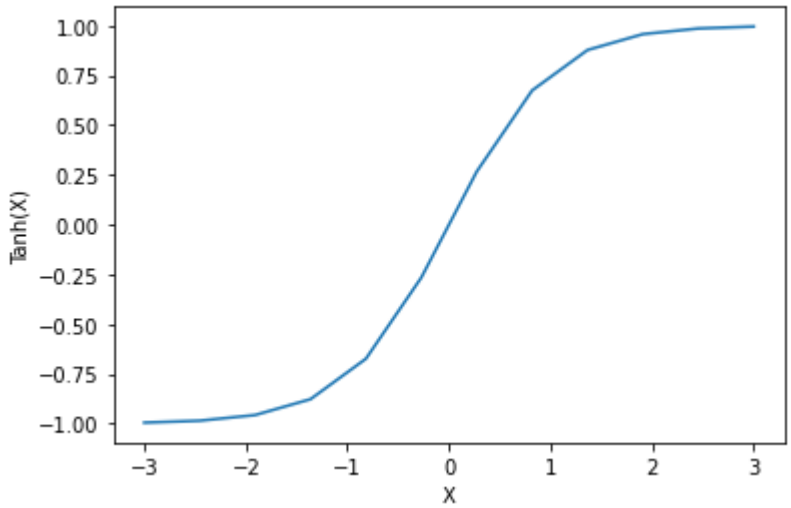


La función de activación de tangente hiperbolica o tanh se define con la siguiente fórmula:

$$Tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \quad (5)$$

La función Tanh se ve representada en la figura 13.

Figura 13. Función de activación Tanh

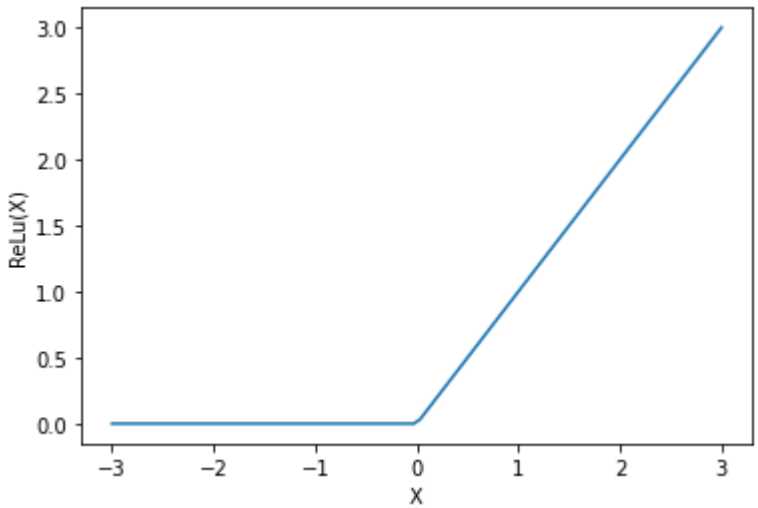


La función de activación lineal rectificada o ReLu se define con la siguiente fórmula:

$$x = \begin{cases} 0 \; si \; x \leq 0 \\ x \; si \; x > 0 \end{cases} \tag{6}$$

La función ReLu se ve representada en la figura 14.

Figura 14. Función de activación ReLu



2.2.4 Recurrent Neural Network

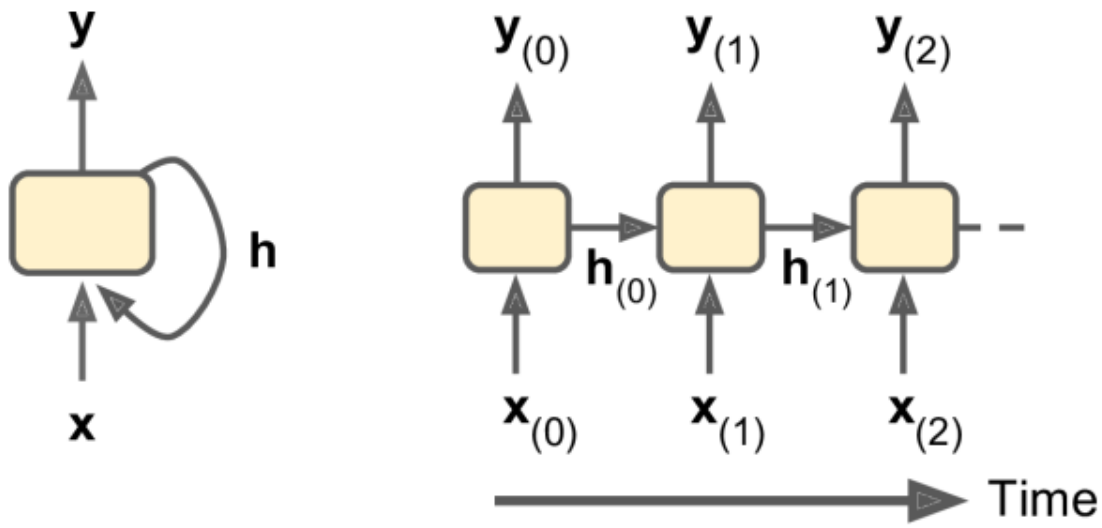
Las redes neuronales recurrentes o *Recurrent Neural Network* (RNN) se diferencian de las redes *feedforward* en que permiten ciclos en el cálculo computacional. Cada ciclo tiene un retraso, por lo que las unidades pueden tomar como entrada un valor calculado a partir de su propia salida en un paso anterior del cálculo. Esto permite que las RNN tengan un estado interno o de memoria, las entradas recibidas en pasos de tiempos anteriores afectan la respuesta de la RNN a la entrada actual. Las RNN pueden ser usadas para realizar cálculos más generales, después de todo son circuitos booleanos con memoria y modelar sistemas neuronales, los cuales tienen conexiones cíclicas. Una de las aplicaciones que se le puede dar a este tipo de redes es para hacer análisis de datos secuenciales, donde se asume la entrada de un valor X en cada paso del tiempo (Stuart & Peter, 2021).

El estado de una celda o neurona de una red recurrente se describe como una función con entradas en el presente paso de tiempo y del paso de tiempo previo, esto se logra expresar en la siguiente formula:

$$h_{(t)} = f(h_{(t-1)}, x_{(t)}) \tag{6}$$

La variable $h_{(t)}$ representa el estado de la celda en el paso de tiempo t (h expresa el estado, también llamado *hidden* u oculto), el valor o salida de esta variable es en base al estado de la celda previo más las entradas actuales (Géron, 2011). El estado oculto es actualizado con la suma de las nuevas entradas en los siguientes pasos de tiempo t, guardando las características con mayor peso y eliminando la información con menor valor, esta secuencia se observa en la figura 15.

Figura 15. Unidad de una red neuronal recurrente a lo largo del paso de tiempo t



Nota. **Adaptado de Géron (2011).**

En el caso del uso de redes recurrentes para secuencias largas, esta es entrenada durante varias etapas, volviéndola una red profunda, lo cual puede ocasionar que se produzcan problemas como la gradiente inestable, haciendo que la etapa de entrenamiento dure bastante tiempo o que el entrenamiento no se logre dar ya que los pesos de las celdas no logran actualizarse. Además, cuando una red recurrente procesa una secuencia larga, gradualmente se olvidará de las primeras entradas que se dieron en la secuencia (Géron, 2011).

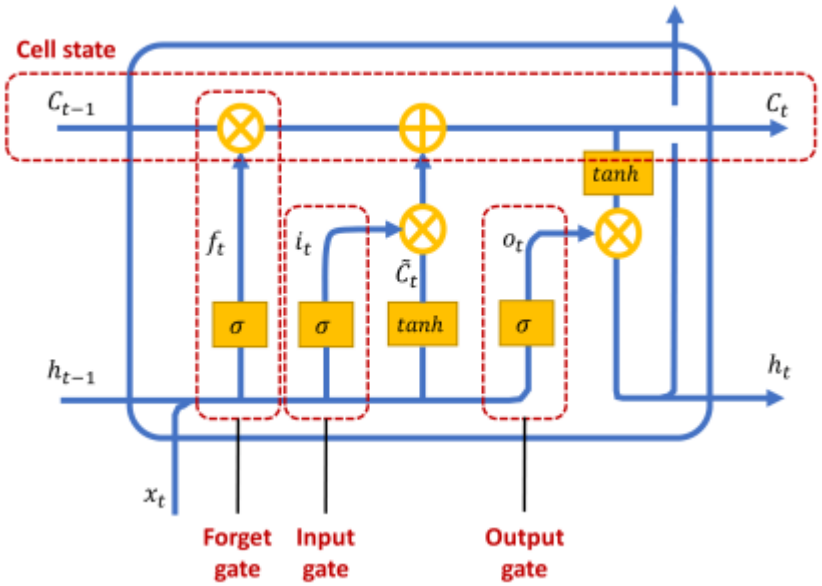
Debido a estos problemas que tienen las redes recurrentes simples, se han desarrollado distintas arquitecturas, las cuales son más complejas y no presentan los problemas mencionados. Una de las principales y más populares son las redes *Long Short Term Memory* o LSTM, que trabajan con extensas secuencias de datos, siendo utilizadas tanto en secuencia de video o palabras.

2.2.5 Long Short Term Memory

Las redes recurrentes LSTM fueron propuestas en 1997 por Sepp Hochreiter and Jürgen Schmidhuber y fue gradualmente mejorada por varios investigadores. El entrenamiento que se tiene con este modelo converge de manera más rápida que una red recurrente simple y a su vez detecta dependencias a largo plazo de los datos de entrenamiento (Géron, 2011).

La clave de los modelos LSTM es el estado de las celdas, añadiendo o agregando información hacia o desde el estado presente de la celda, lo cual se logra a partir de las puertas de cada unidad. Estas puertas están compuestas por una red sigmoide y una operación de multiplicación puntual. La salida de la red sigmoide está entre los valores 0 y 1, donde 0 indica que no se permite el paso de información y 1 indica toda la información. Una celda LSTM contiene tres puertas: puerta de entrada, puerta de salida y puerta de olvido (Yu et al., 2021). En la figura 16 se muestran las partes de una celda LSTM, compuesta por sus 3 puertas, la entrada del estado oculto (*hidden state* o h_{t-1}) y la del estado de celda pasada (C_{t-1}), estas dos últimas se podrían describir como el estado a corto tiempo y estado de largo tiempo respectivamente.

Figura 16. Partes de una celda LSTM



Nota. **Adaptado de Yu et al. (2021).**

Dentro de cada celda, el primer paso es seleccionar el estado de la celda del paso de tiempo anterior y retener parte de la información en el paso de tiempo actual utilizando la puerta de olvido. La puerta de olvido se describe con la ecuación f_t, donde el estado oculto del paso del tiempo anterior h_{t-1} y el valor del paso de tiempo actual x_t se

tienen en cuenta en la función sigmoide σ. Luego de la puerta de olvido, se pasa a la puerta de entrada, esta es comprendida por una capa sigmoide expresada por la ecuación i_t que determina que valores del estado de celda actualizar y por una capa tanh ($\tilde{C}_t$), la cual crea valores intermedios para actualizar el estado de la celda. Por último, está la puerta de salida (o_t) la cual determina los valores de salida de celda con una capa sigmoide y una capa tanh que estandariza los valores del estado de la celda.

$$
\begin{aligned}
f_t &= \sigma(W_f * [h_{t-1}, x_t] + b_f) \\
i_t &= \sigma(W_i * [h_{t-1}, x_t] + b_i) \\
\widetilde{C_t} &= \tanh\,(W_C * [h_{t-1}, x_t] + b_C) \\
C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\
o_t &= \sigma(W_o * [h_{t-1}, x_t] + b_o) \\
h_t &= o_t * \tanh\,(C_t)
\end{aligned}
\tag{7}
$$

Los valores W dentro de las ecuaciones representa las matrices de peso de la celda y los valores b son unos vectores de peso adicional que no requieren entrada y son comúnmente utilizados para cuando se tiene entradas de valor nulo, la salida de la neurona tenga un valor distinto a este.

2.2.6 Series de tiempo

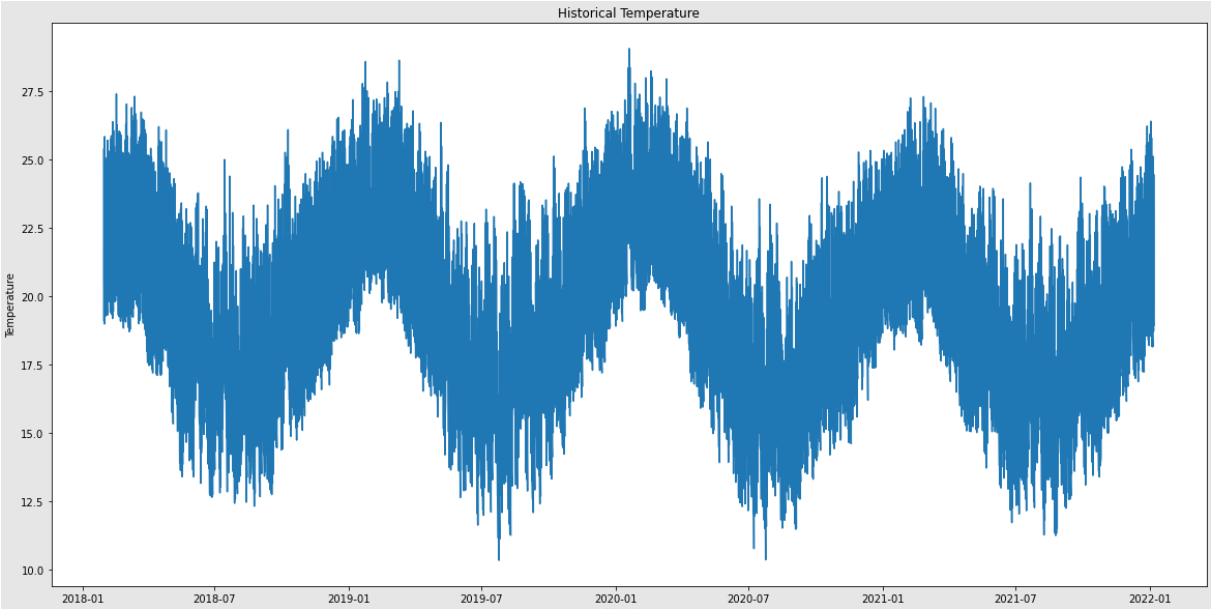
Se puede definir a las series de tiempo como secuencias de datos que toman por índice principal la variable del tiempo, permitiendo visualizar como las variables de un conjunto de datos va modificándose. Existen tanto las series de tiempo con una sola variable y las multivariables, donde dos o más variables modifican su valor respecto del tiempo. El tiempo dentro de las series de tiempo es discreto y es el mismo valor entre cada punto del set de datos (Auffarth, 2021). En la figura 17 se observa un set de datos multivariable.

Figura 17. Serie de tiempo multivariable

Date time	Power	Irradiance	Index Clarity	Zenital Angle	Temperature
2018-01-30 06:00:00	4.59	139.27	0.47	77.76	19.11
2018-01-30 07:00:00	84.45	381.74	0.62	63.88	20.97
2018-01-30 08:00:00	136.36	620.12	0.69	49.84	22.79
2018-01-30 09:00:00	145.30	833.22	0.73	35.73	24.12
2018-01-30 10:00:00	149.76	900.95	0.69	21.66	24.87

A partir del set de datos se puede graficar cada una de las variables y ver su comportamiento histórico respecto al tiempo, observando sus valores máximos y mínimos además de la tendencia o secuencia propia de la variable. Por ejemplo, en la figura 18, se visualiza el intervalo de valores que tiene la variable temperatura dentro de un período de cuatro años, de igual forma se observa una secuencia cíclica a lo largo de cada año.

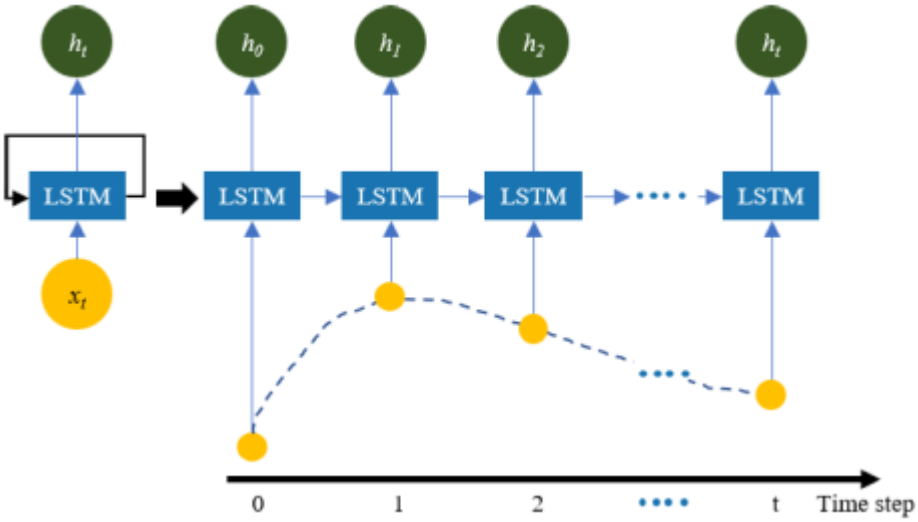
Figura 18. Histórico de variable de temperatura



El aumento del monitoreo y la recopilación de datos trae consigo la necesidad de técnicas estadísticas y de machine Learning aplicadas justamente a las series de tiempo con el objetivo de predecir y caracterizar el comportamiento de sistemas más complejos o componentes dentro de un sistema (Auffarth, 2021).

Un ejemplo de aplicación de LSTM para la predicción de series de tiempo es el trabajo desarrollado por (Liu et al., 2021), el cual recopila los datos de generación eléctrica de dos sistemas fotovoltaicos y genera una predicción de la generación de energía solar con un día de antelación. Dentro de los tipos de datos están los climáticos, la salida de potencia eléctrica del inversor de los paneles solares, medidores de energía y controladores. Para esto, utilizan un modelo LSTM simplificado, el cual logra superar modelos más convencionales como el modelo multicapa de perceptrones. La figura 19 ilustra el ingreso de datos al algoritmo LSTM y a partir de este se tienen valores de salida que generan los valores de predicción.

Figura 19. Entradas y salida de modelo LSTM



Nota. **Adaptado de Liu et al. (2021).**

UNIVERSITAS STUDIORUM PIURENSIS

Capítulo 3

Desarrollo de los Recursos Energéticos Renovables (RER) en el Perú

3.1 Estado de los RER en el Perú

En el Perú, la producción de electricidad creció a una tasa anual de 7.04% en los últimos 21 años: pasó de 12 170 GWh a 50 817 GWh entre 1997 y 2018. La matriz energética del país se ha modificado en las dos últimas décadas. Desde 1997 hasta 2003, la participación de la producción hidroeléctrica fue en promedio de 89.7%. A partir del inicio del Proyecto Camisea en 2004, la producción eléctrica con gas natural se incrementó, llegando a tener una participación máxima de 48% en el 2014. Luego se redujo por el ingreso de las centrales de recursos energéticos renovables (RER), que han pasado de una participación nula en el 2008 a una de 7.2% en el 2018 (OSINERGMIN, 2019).

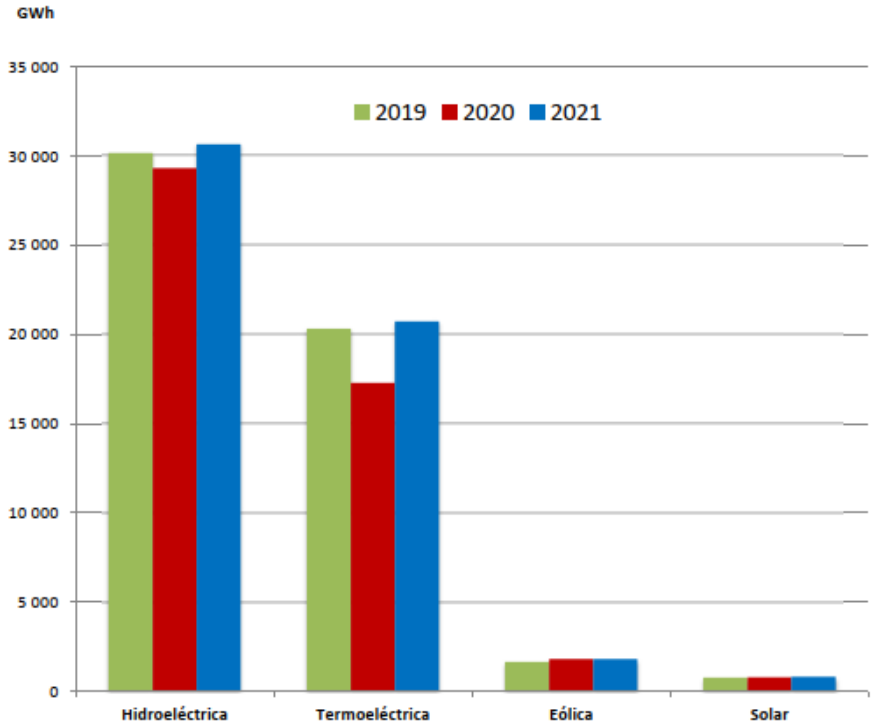
Actualmente el valor de la participación de las RER ha aumentado, para 2021 tuvieron una participación de 5 277.45 GWh, lo cual equivale a una producción del 9.78% del sistema eléctrico interconectado nacional (SEIN).

El total de la producción de energía eléctrica de las empresas generadoras integrantes del Comité de Operación Económica del Sistema Interconectado Nacional (COES) en el año 2021 fue de 50 990.35 GWh, lo cual representa un incremento de 4 803.71 GWh (9.77%) en comparación con el año 2020. La producción de electricidad con centrales hidroeléctricas durante el 2021 fue de 30 664.41 GWh, mientras que la producción eólicas fue de 1800.64 GWh y las centrales solares fue de 801.92 GWh, los cuales tuvieron una participación de 3.34% y 1.49% respectivamente (COES, 2022b).

Estos aportes realizados por la generación eólica y solar tienen un porcentaje de participación bastante reducido a comparación a generaciones como la termoeléctrica, que en el año 2021 fue de 20 723.38 GWh, lo cual representaba un aumento del casi 20% a diciembre del año 2020 o del gas natural de Camisea, que fue de 35.87%.

En la figura 20 se representa gráficamente esta diferencia entre los aportes eléctricos al SEIN entre los años 2019 y 2021.

Figura 20. Producción de energía eléctrica por tipo de generación en el SEIN



Nota. **Adaptado de COES (2022b).**

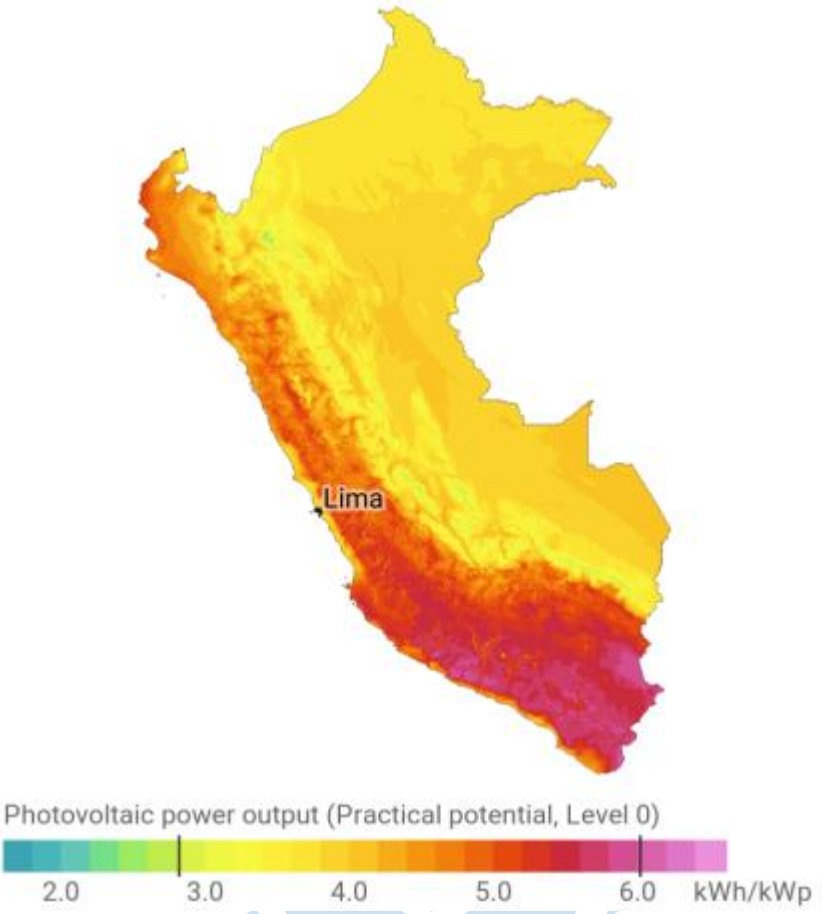
3.1.1 Potencial de generación solar en el Perú

El Perú está ubicado en una zona privilegiada para el desarrollo de la energía solar debido a su proximidad a la línea ecuatorial, lo que permite contar con más horas de sol durante la mayor parte del año con respecto, por ejemplo, a los países nórdicos. Para el desarrollo de proyectos de tecnologías fotovoltaicas, se necesitan análisis de irradiación horizontal global (la cantidad total de radiación de onda corta recibida desde un punto elevado por una superficie horizontal hacia el suelo). Este valor es la suma de la irradiación normal directa más la irradiación difusa horizontal (OSINERGMIN, 2019).

Los departamentos de Arequipa, Moquegua y Tacna son los que presentan mejores condiciones para el desarrollo de tecnologías fotovoltaicas y solar térmica. En estas tres regiones al sur del país, la irradiación directa normal se encuentra en el intervalo 6.8 kWh/m^2 y 7 kWh/m^2, y para la irradiación directa normal se encuentra en el intervalo 7.5 kWh/m^2 y 8.5 kWh/m^2. La información de los mapas de irradiación es esencial para la primera etapa del ciclo de vida de un proyecto de energía solar: prospección y evaluaciones preliminares. Posterior a esto, se requerirán datos de recursos solares y meteorológicos de gran calidad para el desarrollo del proyecto, seguimiento y gestión de activos, y por último, previsión y gestión operativa de plantas de energía solar (OSINERGMIN, 2019).

La figura 21 presenta el mapa del Perú y su irradiación horizontal, siendo bastante prominente en el sur del país.

Figura 21. Mapa de irradiación horizontal del Perú



Nota. **Adaptado de ESMAP (2020).**

3.1.2 Subastas y planificación de RER On-Grid

El desarrollo de las subastas con recursos energéticos renovables (RER) en el Perú tuvo como marco legal el Decreto Legislativo (DL) N° 1002 y su reglamento, Decreto Supremo (DS) N° 050-2008-EM, este último fue actualizado en 2011 mediante el DS N° 012-2011-EM. Según el reglamento, las subastas son un proceso de concurso público convocado y conducido por el Organismo Supervisor de la Inversión en Energía y Minería (Osinergmin). Su finalidad es asignar la tarifa de adjudicación a cada proyecto de generación con RER (solar, eólico, geotérmico, biomasa, hidroeléctrico hasta 20 MW) hasta cubrir la energía requerida en sus bases. De esta manera, se busca promocionar la inversión para la generación de electricidad con el uso de energías renovables, a fin de diversificar la matriz energética. Esto constituye un avance hacia una política de seguridad energética y de protección al ambiente (OSINERGMIN, 2019).

Desde 2009 hasta hoy se han realizado cuatro subastas RER para el SEIN y una subasta RER *off-grid*. Así, se logró avanzar con la diversificación de las fuentes de generación eléctrica en el país. Nueve años después de la primera subasta, casi el 8% del parque generador está compuesto por energías renovables no convencionales. Luego de cuatro subastas se han presentado 249 proyectos, de los cuales se han adjudicado 64. El avance tecnológico y señales adoptadas por el regulador (como no revelar los precios máximos) han ayudado a reducir los precios ofertados (OSINERGMIN, 2019).

A manera de resumen se presentan en la tabla 1 las cuatro subastas on-grid en el Perú, las cuales serán descritas.

Tabla 1. Resumen de subastas on-grid en el Perú

Características relevantes		Primera subasta	Segunda subasta	Tercera subasta	Cuarta subasta
Número de proyectos	Ofertados	56	25	24	131
	Adjudicados	27	10	14	43
Energía (GWh)	Requerida	4380.0	1981.0	1620.0	1750.0
	Ofertada	3810.0	2894.2	1763.7	19 650.6
	Adjudicada	1971.6	1152.7	1277.9	139.2
Solar (USD/MWh)	Promedio Ponderado Ofertado	243.3	186.3	-	57.4
	Mínimo adjudicado	215.0	119.9	-	48.0
	Promedio ponderado adjudicado	221.1	119.9	-	48.1
Eólica (USD/MWh)	Promedio Ponderado Ofertado	89.7	75.4	-	50.5
	Mínimo Adjudicado	65.5	69.9	-	37.8

Características relevantes		Primera subasta	Segunda subasta	Tercera subasta	Cuarta subasta
	Promedio ponderado adjudicado	80.4	69.9	-	40.4
Biomasa (USD/MWh)	Promedio Ponderado Ofertado	63.6	71.3	-	77.0
	Mínimo Adjudicado	52.0	100.0	-	77.0
	Promedio ponderado adjudicado	63.5	100	-	77.0
Hidroeléctrica (USD/MWh)	Promedio Ponderado Ofertado	60.0	55.6	59.6	47.3
	Mínimo Adjudicado	55.0	47.4	50.5	44.0
	Promedio ponderado adjudicado	60.0	53.2	56.4	46.3

Nota. **Adaptado de OSINERGMIN (2019).**

3.1.2.1 Primera subasta. Esta primera subasta constituyó el paso inicial para promocionar los recursos energéticos renovables. Esta subasta se realizó en dos etapas; una convocatoria entre agosto del 2009 y febrero de 2010 y otra entre marzo de 2010 y agosto de 2010, sin embargo, la segunda convocatoria de esta subasta quedó desierta por no existir adjudicatarios a excepción de una. A pesar de esto, el 84% de los proyectos que se presentaron lograron adjudicarse la buena pro. Las fuentes RER que tuvieron mejor desempeño fueron las de biomasa e hidroeléctricas (OSINERGMIN, 2019).

El éxito de la primera convocatoria respondió a los precios competitivos. Las tecnologías estuvieron, en el peor de los casos, 18% por debajo de los precios base. El caso más resaltante fue de biomasa, donde el precio adjudicado fue 47% menor de lo esperado. La menor tarifa fue del proyecto de biomasa Paramonga: 52 USD/MWh (OSINERGMIN, 2019).

Tabla 2. Proyectos de la primera subasta

	Primera convocatoria		Segunda convocatoria	
Tecnología	Presentados	Adjudicados	Presentados	Adjudicados
Biomasa	2	2	5	0
Eólica	6	3	0	0
Solar	6	4	3	0
Hidroeléctrica	17	17	17	1
Total	**31**	**26**	**25**	**1**

Nota. **Adaptado de OSINERGMIN (2019).**

3.1.2.2 Segunda subasta. La segunda subasta se realizó sobre el nuevo reglamento de la ley de promoción de energías renovables. Uno de los cambios más importantes fue la omisión de revelar los precios máximos de la subasta por parte del regulador. La presentación de estos ocurriría solo si la subasta se declara desierta o si alguna oferta supera el monto máximo. Esto se debió a raíz de lo que ocurrió en la segunda convocatoria de la primera subasta, cuando las empresas ajustaron sus precios al alza con respecto a la primera al conocer los montos máximos del concurso (OSINERGMIN, 2019).

Esta segunda subasta tuvo una sola convocatoria. Asimismo, las bases del concurso señalaron que solo podrían participar los proyectos que, a la fecha de la convocatoria, no tenían autorización escrita de conexión al Sistema Eléctrico Interconectado nacional (SEIN) por parte del Comité de Operación Económica (COES). Esto restringió la postulación de centrales que anteriormente habían sido construidas (como en el caso de la primera subasta) (OSINERGMIN, 2019).

Los precios fueron más competitivos en la segunda subasta que en la primera. En general, mientras que el promedio ponderado total de la primera fue de 80.4 USD/MWh, en la segunda fue 62.3 USD/MWh. La tecnología solar fue la que mostró mejores precios en comparativa con las demás tecnologías, sin embargo, la que obtuvo mayores ofertas presentadas y proyectos adjudicados fue la hidroeléctrica (OSINERGMIN, 2019).

Tabla 3. Proyectos de la segunda subasta

Tecnología	Presentados	Adjudicados
Biomasa	2	1
Eólica	6	1
Solar	13	1
Hidroeléctrica	16	7
Total	**37**	**10**

Nota. **Adaptado de OSINERGMIN (2019).**

3.1.2.3 Tercera subasta. Se estableció que la energía anual total a subastar sería 1620 GWh, de los cuales 1300GWh serían para proyectos hidroeléctricos RER y, la diferencia, para los de biomasa. La buena pro se otorgó en diciembre del 2013. La subasta se realizó mediante convocatoria solo para tecnología hidroeléctrica. Se presentaron 24 postores y se adjudicaron 16 proyectos. el inicio del proceso fue el 12 de agosto del 2013. La empresa de generación eléctrica Canchayllo presentó tres proyectos, sin embargo, para dos de ellos incluyó la garantía de fiel cumplimiento y quedaron descalificados. Así, solo se suscribieron los contratos para 14 proyectos (OSINERGMIN, 2019).

Tabla 4. Proyectos hidroeléctricos de la tercera subasta

Precio	USD/MWh	Número de proyectos
Mínimo	50.5	14
Promedio ponderado	56.4	
Máximo	64.8	

Nota. **Adaptado de OSINERGMIN (2019).**

3.1.2.2 Cuarta subasta. Esta última subasta se inició en agosto del 2015 y culminó el 17 de mayo del 2016 con el acta de cierre. Se estableció un requerimiento de 125 GWh/año de energía para residuos forestales, 125 GWh/año de energía para residuos agrícolas, 62 GWh/año de energía para residuos sólidos, 573 GWh/año de energía eólica, 415 GWh/año de energía solar, y 450 GWh/año para proyectos hidroeléctricos. La subasta contó con un total de 13 proyectos adjudicados de los cuales seis se encuentran actualmente en operación (OSINERGMIN, 2019).

Tabla 5. Proyectos de la cuarta subasta

Tecnología	Presentados	Adjudicados
Biomasa	2	2
Eólica	41	3
Solar	61	2
Hidroeléctrica	27	6
Total	**131**	**13**

Nota. **Adaptado de OSINERGMIN (2019).**

Se observaron menores precios ofertados y adjudicados para todas las tecnologías, excepto biomasa. El progreso tecnológico (especialmente en tecnología solar) y la competencia entre postres, ocasionó que se presentaran 131 proyectos con una importante reducción en los precios con respecto a las subastas anteriores (OSINERGMIN, 2019).

Dentro de esta cuarta convocatoria, la empresa Enel Green Power Perú S.A. presentó el proyecto central solar Rubí. El 17 de mayo del 2016 se suscribió el Contrato de Concesión para el Suministro de Electricidad al Sistema Eléctrico Interconectado Nacional (SEIN), en el cual la empresa asumió obligación de construir la central solar Fotovoltaica (CSF) Rubí, con una potencia instalada de 144.48 MW para suministrar la energía adjudicada durante 20 años (OSINERGMIN, 2019).

3.2 Central solar Rubí

La central solar Rubí se encuentra en el Departamento de Moquegua, provincia Mariscal Nieto, a una altura de 1 410 msnm (la ubicación del lugar se observa en la figura 22). Cuenta con una capacidad de 144.48 MW, que se obtiene mediante la instalación de 560 880 módulos fotovoltaicos de 320 W, con un número de módulos en serie de 30 paneles (en la figura 23 hay un vistazo de los paneles instalados en sitio), la estructura que sostiene a los paneles son seguidores horizontales de 1 eje. La energía ofertada por la central es de 415 GWh/año y el precio a ofertar es de 47.98 USD/MWh, el monto de inversión aproximado fue de 165 MM USD, según lo indicado por la concesionaria (OSINERGMIN, 2021).

Figura 22. Ubicación de la central solar Rubí

Nota. **Adaptado de OSINERGMIN (2021).**

Figura 23. Paneles solares de la central fotovoltaica Rubí

Nota. Adaptado de Enel (2018).

Además, cuenta con 164 inversores (la figura 24 muestra la caseta de los inversores), los cuales cuentan con una potencia de 1 025 MVA y la tensión de salida de los inversores es de 400 V en AC, de igual forma cuenta con 41 centros de transformación y la potencia de cada centro es de 3 524 MW, la relación de transformación es de 0.4 /33 KV. La conexión de esta central hacia el SEIN se realizó a través de línea de transmisión de 220 kV S.E. Rubí – S.E. Montalvo, de simple circuito de 21.51 km (OSINERGMIN, 2021).

Figura 24. Caseta de inversores y transformadores

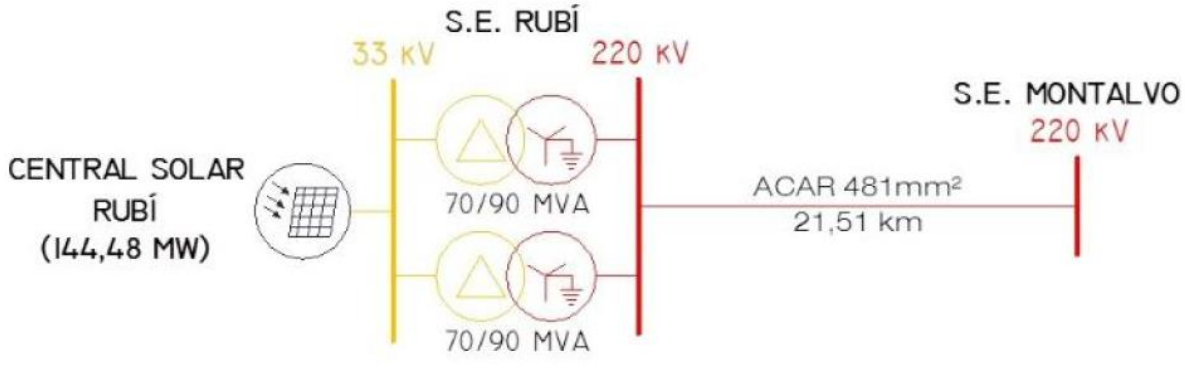
Nota. **Adaptado de OSINERGMIN (2021).**

La central cuenta con su propia subestación eléctrica S.E. Rubí, el cual cuenta con dos transformadores, con una potencial nominal de 70/90 MVA y una relación de transformación de 33/220 KV, ambos son de la marca Chint Electric y su año de fabricación es del 2017 (OSINERGMIN, 2021). En la figura 25 se visualiza el patio de la subestación Rubí y la figura 26 es el diagrama unifilar, este representa la conexión que tiene la central solar para conectarse a la red nacional mediante la S.E. Montalvo.

Figura 25. S.E. Rubí

Nota. **Adaptado por OSINERGMIN (2021).**

Figura 26. Diagrama unifilar



Nota. **Adaptado por OSINERGMIN (2021).**

Dentro del historial del desarrollo del proyecto se pueden destacar los siguientes avances:

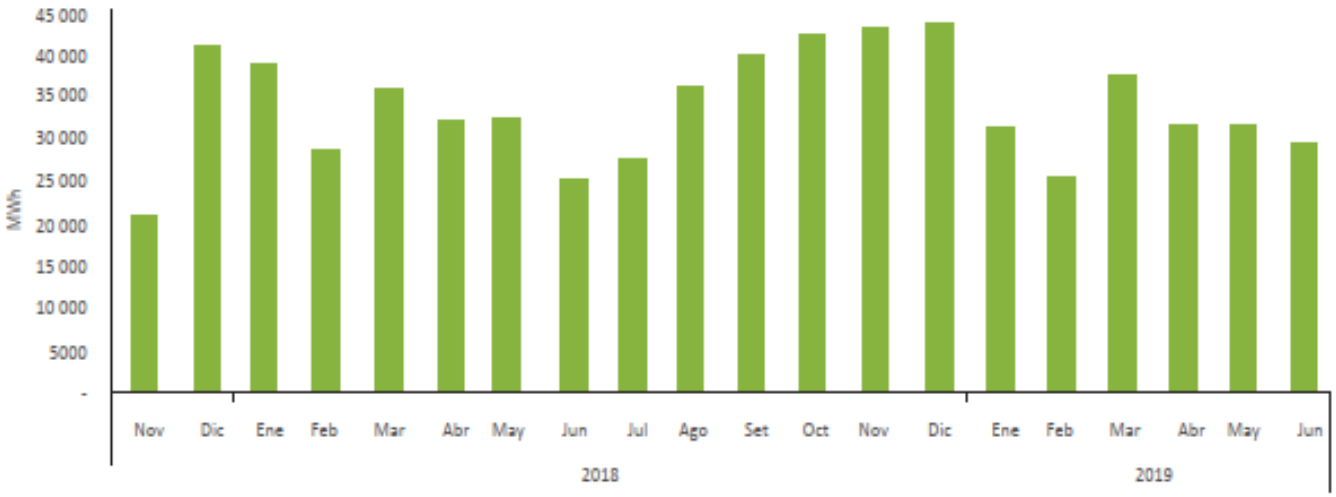
Tabla 6. Avance del desarrollo del proyecto Rubí

Avance	Fecha
La Concesión definitiva para desarrollar la actividad de generación eléctrica en la central de 144.48 MW fue otorgada mediante R.M. N° 328-2017-MEM.	02/08/2017
El COES aprobó el estudio de operatividad de la central.	25/10/2017
El COES autorizó la conexión para las pruebas de puesta en servicio de la central.	31/10/2017
La C.S.F. Rubí se conectó por primera vez con una potencia de 3.5 MW.	10/11/2017
Durante las pruebas se generó 152.6 MW.	29/01/2018
El COES aprobó la puesta en operación comercial de la C.S. Rubí, con una potencia instalada nominal de 144.48 MW, a partir de las 00:00 horas del 30/01/2018.	29/01/2018

Nota. **Adaptado por OSINERGMIN (2021).**

La producción de energía eléctrica por parte de la central solar Rubí se observa en la figura 27, abarcando la generación desde noviembre del 2018 hasta junio del 2019.

Figura 27. Generación eléctrica de CSF Rubí



Nota. Adaptado de OSINERGMIN (2019).

Capítulo 4

Manejo de datos a entrenar

4.1 Teoría de datos

Un set de datos consiste en los datos relacionados con una colección de entidades, con cada entidad descrita en términos de un conjunto de atributos. En su forma más básica, un set de datos se organiza en una matriz de datos de n*m denominada registro analítico, donde n es el número de entidades (filas) y m es el número de atributos (columnas) (Kelleher & Tierney, 2018).

Independiente del tipo de información que se tenga (numérica, vídeo, fotos, textos) esta puede ser agrupada dentro de una estructura o set de datos, a partir de la cual se puede extraer información haciendo uso de algoritmos. Para ingresar la información a los algoritmos y obtener buenos resultados, el set de datos debe pasar un proceso de preparación. A continuación, se describirá las fases para la preparación de un correcto set de datos.

4.1.1 Recopilación de datos

En esta fase se puede dar el caso en que se deba realizar el diseño de recolección de datos por cuenta propia, aunque es posible que la data ya se encuentre almacenada y agrupada, pudiendo estar dentro de unos archivos simples de texto o tablas de base de datos. Este set de datos puede ser de libre uso o ser propiedad de una organización, sin embargo, la cantidad de datos de libre acceso es bastante amplia (Cielen et al., 2016).

En general, existen dos términos para los datos sin procesar recopilados: datos capturados y datos de escape. Los datos capturados recopilan a través de una medición u observación directa que está diseñada para recopilar datos. Por el contrario, los datos de escape son un subproducto de un proceso cuyo propósito principal es algo más que la captura de datos (Kelleher & Tierney, 2018).

Durante la recopilación, se puede verificar si los datos de ingreso son similares a los del documento fuente de donde se les está obteniendo y si los tipos de datos son los correctos, si bien en fases posteriores se realizará a más detalle el análisis de los datos obtenidos, realizarlo en esta fase permite saber si hay errores en la entrada de datos (Cielen et al., 2016).

4.1.2 Limpieza de datos

El propósito de esta fase es limpiar y preparar el set de datos para ser usado en el modelo, por lo cual es bastante importante realizar una correcta preparación ya que de esto depende el tipo de resultados que se obtendrán del modelo, malos datos de entrada son equivalentes a malas salidas del modelo. El modelo requiere formatos específicos de entrada, por lo que la transformación de los tipos de datos debe realizarse en esta etapa (Cielen et al., 2016).

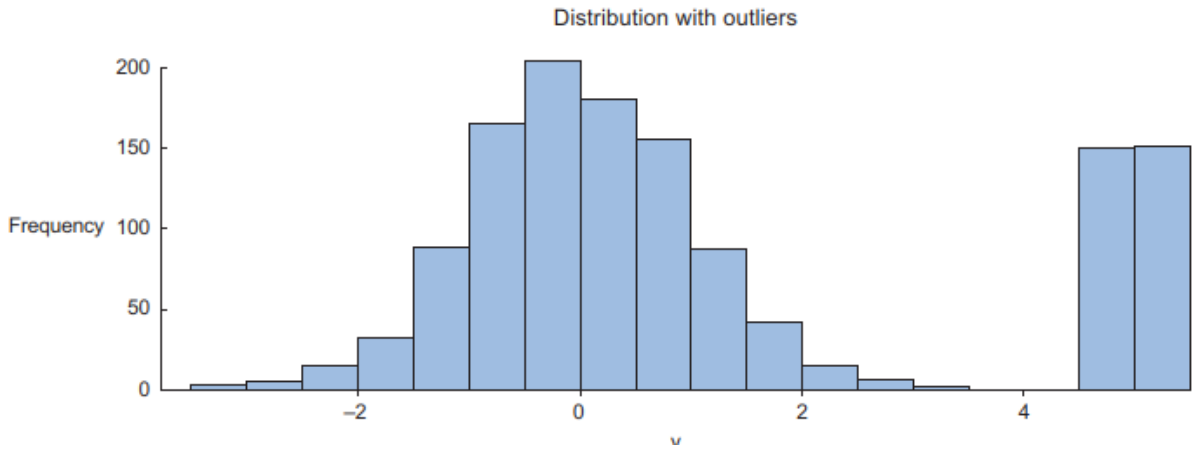
La limpieza se encarga de la eliminación de errores en la data, de tal manera que refleja de manera más real y consistente la representación del proceso del cual se originan estos datos. Hay dos tipos de errores, el primero es el error de interpretación, cuando se da por sentado el valor de tus datos, como decir que la edad de una persona es mayor a 300 años. El segundo tipo de error apunta a inconsistencias entre las fuentes de datos o contra los valores estandarizados, un ejemplo de esto es referenciar un listado de datos bajo distintos nombres o modificar el tipo de datos, lo cual genera una confusión a la hora de extraer y trabajar con este set de datos (Cielen et al., 2016).

Se puede dar el caso que en la categorización de un listado de datos haya un error tipográfico, lo cual puede generar errores en el conteo de resultados o generar categorías innecesarias a las que ya se tienen. Cuando son un conjunto de datos pequeños se puede realizar una revisión total, sin embargo, en el caso de tener grandes cantidades se debe recurrir a herramientas de manejo de datos que analicen los datos y a partir de esto generen un resumen y descripción de todo el set de datos, y en base a esto se pueda analizarlo (Cielen et al., 2016).

De igual forma, a parte de los errores tipográficos están los espacios en blanco dentro de un listado, esto puede perjudicar el performance del modelo, por lo que es preciso ubicarlo dentro de la estructura de datos y partir de esto tener un plan de acción, hay distintas soluciones que se pueden realizar, desde simplemente eliminar todos los valores nulos o reemplazarlo con el valor promedio, esto va a depender del tipo de datos del cual se esté tratando y el uso que se le da.

Otros casos que se pueden dar dentro de un set de datos son los valores atípicos. Un valor atípico se diferencia a que su valor es distante o distinto al resto de datos. La mejor manera de encontrar estos valores es utilizar una gráfica que permita observar los valores máximos y mínimos, un ejemplo de esto se observa en la figura 28.

Figura 28. Distribución de frecuencia de valores

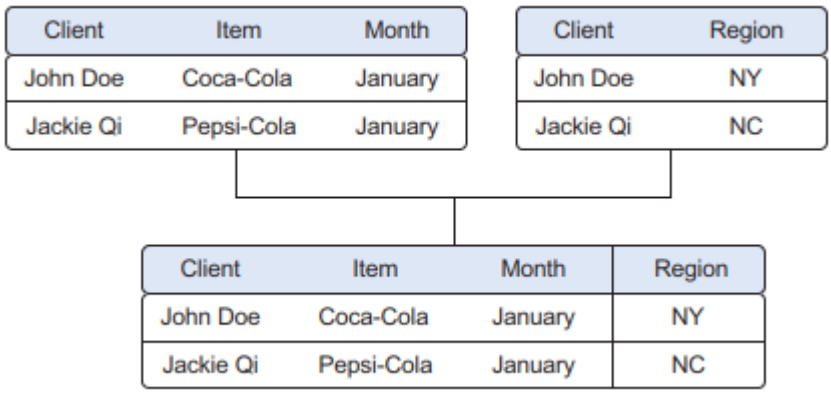


Nota. **Adaptado de Cielen et al. (2016).**

Dependiendo del objetivo final del estudio, estructuras de datos distintas pueden ser unidas y generar nuevas, las cuales serán usadas como información de entrada al modelo. Para realizar esta unión, se puede usar tanto herramientas como Excel o en programas de manejo de estructura de datos como SQL, el cual permite mantener las estructuras iniciales y generar una adicional.

Un ejemplo de esta operación es el de la figura 29, en el cual une dos grupos de datos y genera uno nuevo.

Figura 29. Unión de dos sets de datos

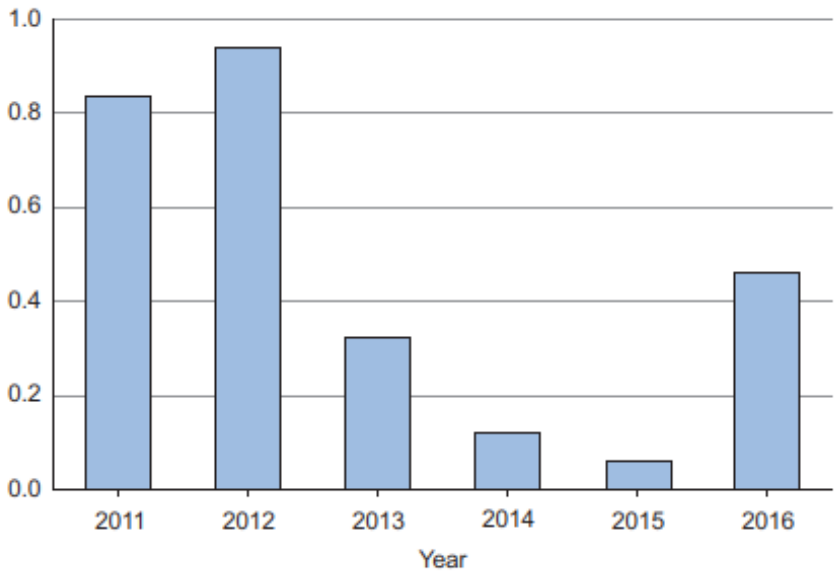


Nota. **Adaptado de Cielen et al. (2016).**

4.1.3 Análisis de exploración de datos

Durante esta fase, se realiza un análisis exploratorio de datos bastante detallado. La información se vuelve más fácil de comprender a partir del uso de gráficos, por lo que se utiliza principalmente técnicas gráficas para comprender los datos y las interacciones entre las variables. Si bien el objetivo de limpiar datos se realizó en la fase previa, es posible que se descubran anomalías que anteriormente no se hallaron, por lo que se puede eliminar o corregirlas. Las técnicas de visualización que pueden ser usadas son gráficos lineales o histogramas (Cielen et al., 2016). La figura 30 es un histograma de un set de datos que se divide en años.

Figura 30. Histograma para análisis de datos.



Nota. **Adaptado de Cielen et al. (2016).**

Las técnicas descritas son principalmente visuales, pero en la práctica ciertamente no se limitan a las técnicas de visualización. La tabulación, el agrupamiento y otras técnicas de modelado también pueden ser parte del análisis exploratorio. Incluso la construcción de modelos simples puede ser parte de esta fase (Cielen et al., 2016).

4.1.4 Construcción del modelo

En esta última fase se implementa y construye el modelo a encargarse de recibir todo el set de datos que se preparó, limpió y analizó en las fases previas con el fin de cumplir con el objetivo trazado inicialmente en el estudio, ya sea una predicción de los siguientes valores en el tiempo o la categorización de la propia data a ingresar.

La construcción de un modelo es un proceso iterativo. La forma en que se construya el modelo depende el tipo de solución que se seleccione, pudiendo aplicar desde estadística clásica o procesos de aprendizaje inteligente. Los pasos a seguir durante el desarrollo del modelo son: selección del modelo y las variables a ingresar, ejecución del modelo y, diagnóstico y comparativa de los resultados (Cielen et al., 2016).

4.2 Datos del COES

El Comité de Operación Económica del Sistema Interconectado Nacional (COES) es una entidad privada que está conformada por los agentes del SEIN (generadores, distribuidores, transmisores y usuarios libre) y sus decisiones son de cumplimiento obligatorio por los agentes. Su finalidad es coordinar la operación de corto, mediano y largo plazo del SEIN al mínimo costo, preservando la seguridad del sistema, el mejor aprovechamiento de los recursos energéticos, así como planificar el desarrollo de la transmisión del SEIN y administrar el mercado al corto plazo (COES, 2022).

El COES dentro de sus actividades, sube a su página web la información referente a la demanda de los agentes del SEIN, reportes, evaluaciones, fallas, entre muchas otras. Uno de los documentos que publica es el Informe de Evaluación de la Operación Diaria (IEOD), donde reporta la generación del día dividiéndola por zona del Perú o tipo de generación. Este informe tiene una sección específica para la generación eléctrica de las centrales RER, agrupando las centrales por empresa propietaria y a su vez detallando la potencia eléctrica de cada central en intervalos de tiempo de media hora, como se observa en la figura 31.

Figura 31. Potencia eléctrica de centrales de Enel Green Power Perú

COES

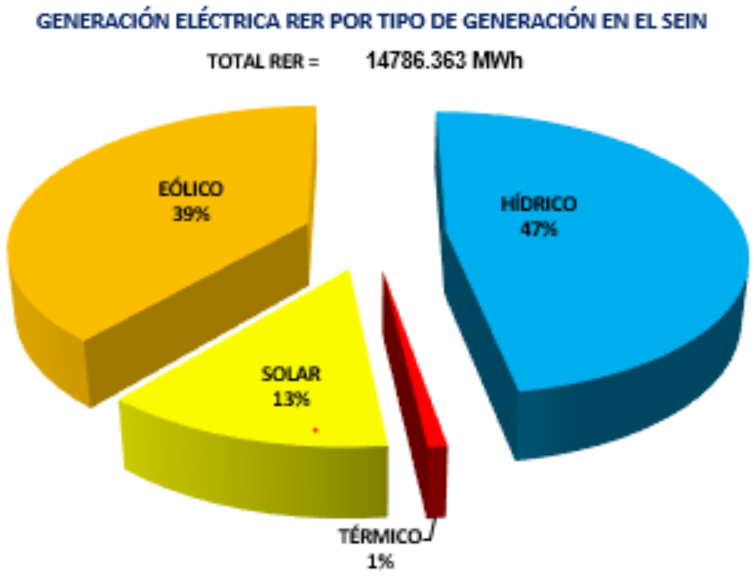
martes, 4 de Mayo de 2021

GENERACIÓN ELÉCTRICA DE LAS CENTRALES RER (MW)		
HORA	ENEL GREEN POWER PERU S.A	
	RUBÍ	WAYRA I
06:00	0.20	56.90
06:30	10.70	39.60
07:00	46.80	38.00
07:30	88.00	34.20
08:00	112.50	38.90
08:30	116.90	31.60
09:00	121.20	30.60
09:30	116.40	45.10
10:00	119.60	62.20
10:30	119.10	80.40
11:00	120.00	88.80
11:30	119.20	91.20
12:00	121.20	101.10

Nota. **Adaptado de COES (2022a).**

El porcentaje de participación diaria de cada tipo de RER se refleja en la figura 32. Esta se verá afectada por las variables ambientales de cada etapa del año.

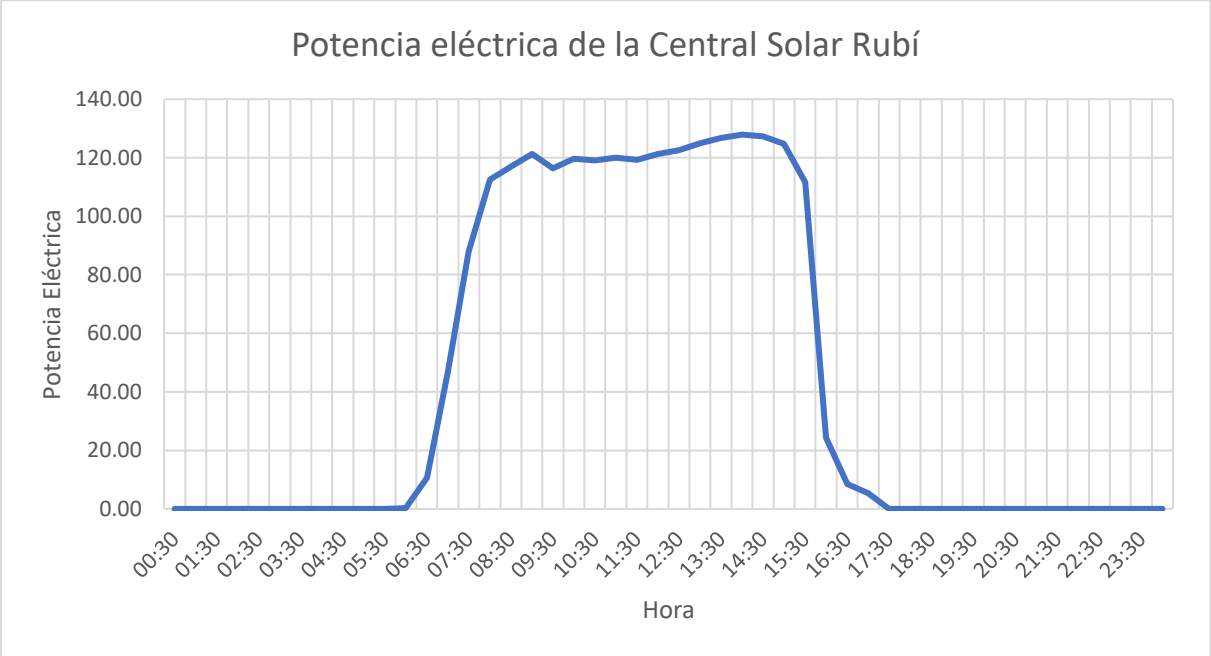
Figura 32. Porcentaje de generación eléctrica por tipo de RER



Nota. **Adaptado de COES (2022a).**

A partir de este informe se obtiene la potencia eléctrica diaria de la central solar Rubí la cual será utilizada como input para el algoritmo. Un ejemplo de la potencia diaria se encuentra en la figura 33, representando la potencia eléctrica de la central solar Rubí el martes 4 de mayo del 2021.

Figura 33. Potencia eléctrica de la central solar Rubí



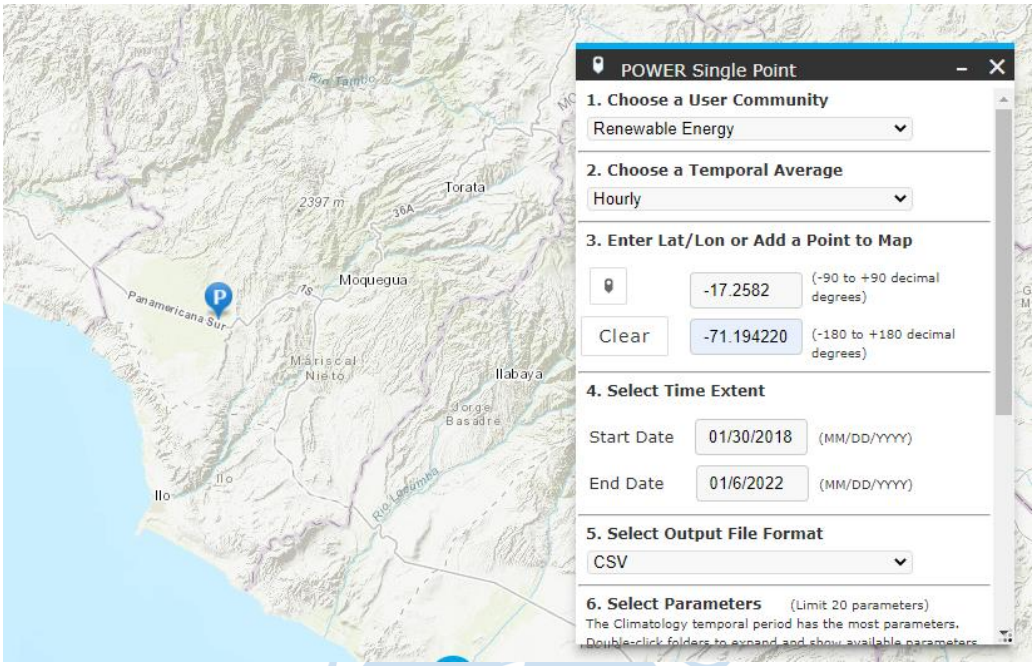
Nota. **Adaptado de COES (2022a).**

4.3 Datos de la NASA

La Administración Nacional de Aeronáutica y el Espacio (NASA), es una agencia estadounidense responsable de la ciencia y tecnología relacionada por el aire y el espacio (NASA, 2022). Dentro de sus aplicaciones de libre acceso, se encuentra su página web *The POWER Project*, el cual permite tener acceso a datos climatológicos de las coordenadas a la cual se ingrese, y se puede seleccionar el período de fechas que se desea obtener y el intervalo de tiempo que se desea tener los datos, siendo lo mínimo el intervalo horario.

Una vez ingresadas las coordenadas de la central solar Rubí, la ubicación es fijada dentro del mapa de la página, tal como se observa en la figura 34. Luego de la selección de la ubicación, se seleccionan los parámetros que se desean obtener de lugar y el tipo de archivo a generar, en este caso se selecciona CSV.

Figura 34. Toma de datos meteorológicos de la central solar Rubí



Nota. **Adaptado de *The POWER Project* (2022).**

Los parámetros seleccionados de este repositorio son los siguientes:

- Irradiación solar
- Índice de claridad del cielo
- Ángulo cenital solar
- Temperatura del aire a dos metros encima de la superficie terrestre.

Estas cuatro variables influyen en la generación de potencia eléctrica de la central solar, por lo cual junto con los datos obtenidos del COES forman parte del set de datos a utilizarse como *input* para el algoritmo LSTM.

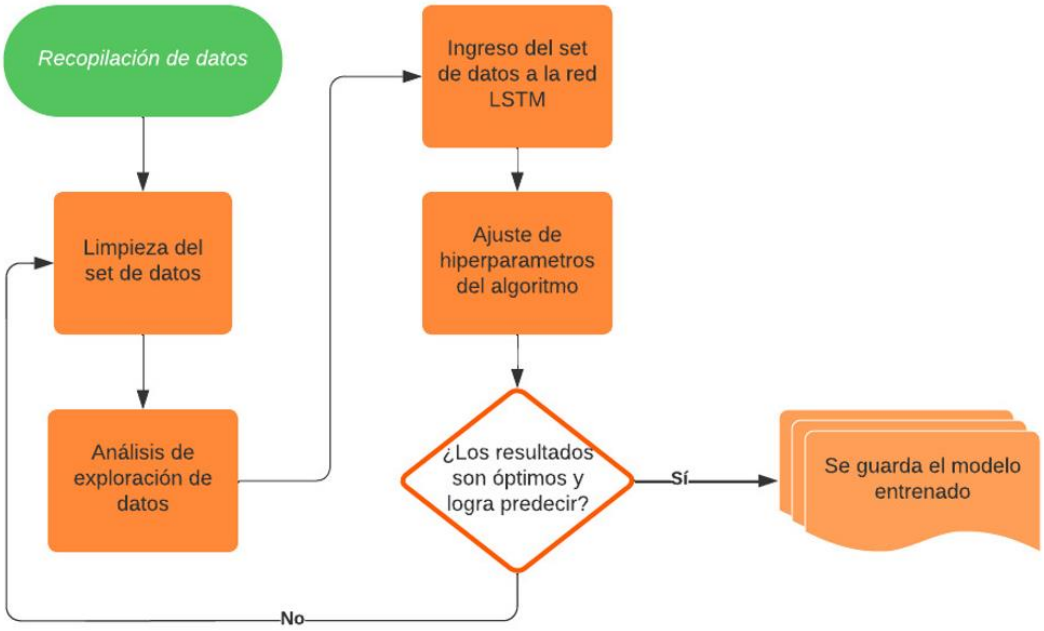
Capítulo 5

Implementación de la red neuronal LSTM

A partir de la recopilación de los datos obtenidos tanto de los organismos del COES como de la NASA; abordados en el capítulo anterior, se procederá a describir el proceso a realizar para obtener un correcto set de datos listo para ser utilizado como datos de entrada por el algoritmo y la implementación de la red neuronal LSTM, con la finalidad de obtener un óptimo modelo predictivo.

En vista de una mayor claridad de este proceso a realizar, la figura 35 presenta el diagrama de flujo donde se observa el desarrollo tanto del manejo de datos como la implementación de la red neuronal y sus ajustes. Si los resultados no son los esperados o se refleja que no se da una correcta lectura, se vuelve al primer paso con el fin de eliminar cualquier error no identificado de forma inicial.

Figura 35. Diagrama de flujo del manejo de datos e implementación de la red neuronal



5.1 Preprocesamiento de los datos a la red neuronal

La toma de datos se realiza desde finales de enero del 2018 hasta inicios de enero del 2022, ambos *sets* de datos se tienen en formato Excel, uno de potencia eléctrica y otro de estado meteorológico, ambos deben ser unidos para crear el *set* de datos final a ingresar a la red neuronal. A continuación, se presenta en las figuras 36 y 37 los *sets* de datos.

Figura 36. Set de datos meteorológicos

YEAR	MO	DY	HR	ALLSKY_SFC_SW_DWN	ALLSKY_KT	SZA	T2M
2018	1	30	5	8.23	0.32	88.94	18.21
2018	1	30	6	139.27	0.47	77.76	19.11
2018	1	30	7	381.74	0.62	63.88	20.97
2018	1	30	8	620.12	0.69	49.84	22.79
2018	1	30	9	833.22	0.73	35.73	24.12
2018	1	30	10	900.95	0.69	21.66	24.87
2018	1	30	11	982.41	0.71	8.12	25.26
2018	1	30	12	932.15	0.67	8.4	25.4
2018	1	30	13	944.7	0.73	21.98	25.23
2018	1	30	14	813.91	0.72	36.06	24.8
2018	1	30	15	466.18	0.52	50.17	24.11
2018	1	30	16	269.32	0.45	64.21	23.08
2018	1	30	17	100.8	0.35	78.1	21.84
2018	1	30	18	8.27	0.37	89.08	20.67
2018	1	30	19	0	-999	-999	19.99

Figura 37. Set de datos de potencia eléctrica

Date	Day	Month	Year	Hour	Date time	Power
01/30/2018	30	1	2018	00:30	01/30/2018 00:30	0.00
01/30/2018	30	1	2018	01:00	01/30/2018 01:00	0.00
01/30/2018	30	1	2018	01:30	01/30/2018 01:30	0.00
01/30/2018	30	1	2018	02:00	01/30/2018 02:00	0.00
01/30/2018	30	1	2018	02:30	01/30/2018 02:30	0.00
01/30/2018	30	1	2018	03:00	01/30/2018 03:00	0.00
01/30/2018	30	1	2018	03:30	01/30/2018 03:30	0.00
01/30/2018	30	1	2018	04:00	01/30/2018 04:00	0.00
01/30/2018	30	1	2018	04:30	01/30/2018 04:30	0.00
01/30/2018	30	1	2018	05:00	01/30/2018 05:00	0.00
01/30/2018	30	1	2018	05:30	01/30/2018 05:30	0.00
01/30/2018	30	1	2018	06:00	01/30/2018 06:00	4.59
01/30/2018	30	1	2018	06:30	01/30/2018 06:30	38.79
01/30/2018	30	1	2018	07:00	01/30/2018 07:00	84.45
01/30/2018	30	1	2018	07:30	01/30/2018 07:30	123.53
01/30/2018	30	1	2018	08:00	01/30/2018 08:00	136.36
01/30/2018	30	1	2018	08:30	01/30/2018 08:30	138.59
01/30/2018	30	1	2018	09:00	01/30/2018 09:00	145.30

Tal y como se observa, tanto en el set de los datos meteorológicos como el de potencia eléctrica se encuentra a detalle la fecha y hora de cada toma de datos, y en las columnas siguientes, las variables a utilizar.

El intervalo de tiempo entre ambos *sets* es distinto, siendo el del COES media hora y el de NASA de una hora, por lo que se procede a modificar el intervalo del COES y unirlos. El resultado de esta unión se visualiza en la figura 38.

Figura 38. Unión de sets de datos

Date	Day	Mont	Year	Hour	Date time	Powe	Irradiance	Index Clarity	Zenital Angle	Temperature
01/30/2018	30	1	2018	01:00	01/30/2018 01:00	0.00	0.00	-999	-999	18.73
01/30/2018	30	1	2018	02:00	01/30/2018 02:00	0.00	0	-999	-999	18.58
01/30/2018	30	1	2018	03:00	01/30/2018 03:00	0.00	0	-999	-999	18.44
01/30/2018	30	1	2018	04:00	01/30/2018 04:00	0.00	0	-999	-999	18.32
01/30/2018	30	1	2018	05:00	01/30/2018 05:00	0.00	8.23	0.32	88.94	18.21
01/30/2018	30	1	2018	06:00	01/30/2018 06:00	4.59	139.27	0.47	77.76	19.11
01/30/2018	30	1	2018	07:00	01/30/2018 07:00	84.45	381.74	0.62	63.88	20.97
01/30/2018	30	1	2018	08:00	01/30/2018 08:00	136.36	620.12	0.69	49.84	22.79
01/30/2018	30	1	2018	09:00	01/30/2018 09:00	145.30	833.22	0.73	35.73	24.12
01/30/2018	30	1	2018	10:00	01/30/2018 10:00	149.76	900.95	0.69	21.66	24.87
01/30/2018	30	1	2018	11:00	01/30/2018 11:00	117.26	982.41	0.71	8.12	25.26
01/30/2018	30	1	2018	12:00	01/30/2018 12:00	148.12	932.15	0.67	8.4	25.4
01/30/2018	30	1	2018	13:00	01/30/2018 13:00	136.34	944.7	0.73	21.98	25.23
01/30/2018	30	1	2018	14:00	01/30/2018 14:00	81.89	813.91	0.72	36.06	24.8
01/30/2018	30	1	2018	15:00	01/30/2018 15:00	104.28	466.18	0.52	50.17	24.11
01/30/2018	30	1	2018	16:00	01/30/2018 16:00	71.78	269.32	0.45	64.21	23.08
01/30/2018	30	1	2018	17:00	01/30/2018 17:00	58.40	100.8	0.35	78.1	21.84
01/30/2018	30	1	2018	18:00	01/30/2018 18:00	7.00	8.27	0.37	89.08	20.67
01/30/2018	30	1	2018	19:00	01/30/2018 19:00	0.00	0	-999	-999	19.99

Analizando el set de datos, se observa que hay presencia del valor -999 en las variables meteorológicas, este valor difiere del rango válido para estas variables. El uso de este número es para representar los valores faltantes o que están fuera del rango de disponibilidad. Además, este valor se encuentra finalizando la tarde hasta las últimas horas de la noche. La razón por la cual sólo está presente dentro de este rango horario se debe a que la irradiación, índice de claridad y ángulo cenital dependen de la presencia del sol; por lo que, dependiendo de la etapa del año, la presencia de este valor estará presente a una mayor o menor cantidad de horas del día. La toma de datos de temperatura no se ve afectada por la ausencia del sol y sus valores se encuentran dentro de lo esperado.

Con el fin de tener un correcto intervalo y un mejor manejo de datos, todos los valores de -999 son reemplazados por el valor de 0, quedando como en la figura 39.

Figura 39. Reemplazo de valores faltantes o no medidos a cero

Date	Day	Mont	Year	Hour	Date time	Powe	Irradiance	Index Clarity	Zenital Angle	Temperature
01/30/2018	30	1	2018	01:00	01/30/2018 01:00	0.00	0.00	0	0	18.73
01/30/2018	30	1	2018	02:00	01/30/2018 02:00	0.00	0	0	0	18.58
01/30/2018	30	1	2018	03:00	01/30/2018 03:00	0.00	0	0	0	18.44
01/30/2018	30	1	2018	04:00	01/30/2018 04:00	0.00	0	0	0	18.32
01/30/2018	30	1	2018	05:00	01/30/2018 05:00	0.00	8.23	0.32	88.94	18.21
01/30/2018	30	1	2018	06:00	01/30/2018 06:00	4.59	139.27	0.47	77.76	19.11
01/30/2018	30	1	2018	07:00	01/30/2018 07:00	84.45	381.74	0.62	63.88	20.97
01/30/2018	30	1	2018	08:00	01/30/2018 08:00	136.36	620.12	0.69	49.84	22.79
01/30/2018	30	1	2018	09:00	01/30/2018 09:00	145.30	833.22	0.73	35.73	24.12
01/30/2018	30	1	2018	10:00	01/30/2018 10:00	149.76	900.95	0.69	21.66	24.87
01/30/2018	30	1	2018	11:00	01/30/2018 11:00	117.26	982.41	0.71	8.12	25.26
01/30/2018	30	1	2018	12:00	01/30/2018 12:00	148.12	932.15	0.67	8.4	25.4
01/30/2018	30	1	2018	13:00	01/30/2018 13:00	136.34	944.7	0.73	21.98	25.23
01/30/2018	30	1	2018	14:00	01/30/2018 14:00	81.89	813.91	0.72	36.06	24.8
01/30/2018	30	1	2018	15:00	01/30/2018 15:00	104.28	466.18	0.52	50.17	24.11
01/30/2018	30	1	2018	16:00	01/30/2018 16:00	71.78	269.32	0.45	64.21	23.08
01/30/2018	30	1	2018	17:00	01/30/2018 17:00	58.40	100.8	0.35	78.1	21.84
01/30/2018	30	1	2018	18:00	01/30/2018 18:00	7.00	8.27	0.37	89.08	20.67
01/30/2018	30	1	2018	19:00	01/30/2018 19:00	0.00	0	0	0	19.99

Debido a que los paneles solares sólo producen su energía eléctrica cuando están expuestas al sol y durante la noche su generación es cero, por lo cual no habría valores a predecir durante esta mitad del día, se prevé trabajar sólo con las horas en las cuales la central solar entre en operación. Este horario de operatividad va en promedio desde las seis de la mañana hasta las seis de la tarde, lo cual representan 12 horas de generación eléctrica por parte de la central solar Rubí. El resultado se encuentra en la figura 40.

Figura 40. Set de datos con sólo horas de generación eléctrica

Date	Day	Month	Year	Hour	Date time	Power	Irradiance	Index Clarity	Zenital Angle	Temperature
01/30/2018	30	1	2018	06:00	01/30/2018 06:00	4.59	139.27	0.47	77.76	19.11
01/30/2018	30	1	2018	07:00	01/30/2018 07:00	84.45	381.74	0.62	63.88	20.97
01/30/2018	30	1	2018	08:00	01/30/2018 08:00	136.36	620.12	0.69	49.84	22.79
01/30/2018	30	1	2018	09:00	01/30/2018 09:00	145.30	833.22	0.73	35.73	24.12
01/30/2018	30	1	2018	10:00	01/30/2018 10:00	149.76	900.95	0.69	21.66	24.87
01/30/2018	30	1	2018	11:00	01/30/2018 11:00	117.26	982.41	0.71	8.12	25.26
01/30/2018	30	1	2018	12:00	01/30/2018 12:00	148.12	932.15	0.67	8.4	25.4
01/30/2018	30	1	2018	13:00	01/30/2018 13:00	136.34	944.7	0.73	21.98	25.23
01/30/2018	30	1	2018	14:00	01/30/2018 14:00	81.89	813.91	0.72	36.06	24.8
01/30/2018	30	1	2018	15:00	01/30/2018 15:00	104.28	466.18	0.52	50.17	24.11
01/30/2018	30	1	2018	16:00	01/30/2018 16:00	71.78	269.32	0.45	64.21	23.08
01/30/2018	30	1	2018	17:00	01/30/2018 17:00	58.40	100.8	0.35	78.1	21.84
01/30/2018	30	1	2018	18:00	01/30/2018 18:00	7.00	8.27	0.37	89.08	20.67
01/31/2018	31	1	2018	06:00	01/31/2018 06:00	9.09	111.7	0.38	77.87	18.98
01/31/2018	31	1	2018	07:00	01/31/2018 07:00	69.40	333.77	0.54	63.96	20.43
01/31/2018	31	1	2018	08:00	01/31/2018 08:00	120.73	544.68	0.6	49.91	22.26
01/31/2018	31	1	2018	09:00	01/31/2018 09:00	144.58	787.23	0.7	35.8	23.68
01/31/2018	31	1	2018	10:00	01/31/2018 10:00	140.58	916.55	0.7	21.71	24.74
01/31/2018	31	1	2018	11:00	01/31/2018 11:00	150.85	900.3	0.65	8.16	25.51

Luego de haber realizado esta modificación, se procede a pasar el set de datos a un formato CSV, el cual es utilizado comúnmente para realizar trabajos de datos con las librerías de Python, evitando problemas de compatibilidad en la lectura de datos. Asimismo, la variable fecha es verificada, ya que, si esta no se encuentra en el formato esperado, se tendría una lectura incorrecta, ocasionando errores posteriores. Por lo que el formato del data set queda como en la figura 41.

Figura 41. Set de datos en formato CSV

```
Date,Day,Month,Year,Hour,Date time,Power,Irradiance,Index Clarity,Zenital Angle,Temperature
01/30/2018,30,1,2018,6:00,01/30/2018 06:00,4.59,139.27,0.47,77.76,19.11
01/30/2018,30,1,2018,7:00,01/30/2018 07:00,84.45,381.74,0.62,63.88,20.97
01/30/2018,30,1,2018,8:00,01/30/2018 08:00,136.36,620.12,0.69,49.84,22.79
01/30/2018,30,1,2018,9:00,01/30/2018 09:00,145.30,833.22,0.73,35.73,24.12
01/30/2018,30,1,2018,10:00,01/30/2018 10:00,149.76,900.95,0.69,21.66,24.87
01/30/2018,30,1,2018,11:00,01/30/2018 11:00,117.26,982.41,0.71,8.12,25.26
01/30/2018,30,1,2018,12:00,01/30/2018 12:00,148.12,932.15,0.67,8.4,25.4
01/30/2018,30,1,2018,13:00,01/30/2018 13:00,136.34,944.7,0.73,21.98,25.23
01/30/2018,30,1,2018,14:00,01/30/2018 14:00,81.89,813.91,0.72,36.06,24.8
01/30/2018,30,1,2018,15:00,01/30/2018 15:00,104.28,466.18,0.52,50.17,24.11
01/30/2018,30,1,2018,16:00,01/30/2018 16:00,71.78,269.32,0.45,64.21,23.08
01/30/2018,30,1,2018,17:00,01/30/2018 17:00,58.40,100.8,0.35,78.1,21.84
01/30/2018,30,1,2018,18:00,01/30/2018 18:00,7.00,8.27,0.37,89.08,20.67
```

El archivo CSV es cargado a la plataforma Google Colab, el cual permite trabajar con el lenguaje de programación Python. Desde un inicio del proceso, se importan las librerías numpy, pandas y matplotlib, las cuales permiten realizar análisis y manipulación de datos con mayor facilidad; por ejemplo, la lectura y visualización del set de datos con formato CSV se realiza con la librería pandas, esto se refleja en la figura 42.

Figura 42. Set de datos en Colab

Date time	Power	Irradiance	Index Clarity	Zenital Angle	Temperature
2018-01-30 06:00:00	4.59	139.27	0.47	77.76	19.11
2018-01-30 07:00:00	84.45	381.74	0.62	63.88	20.97
2018-01-30 08:00:00	136.36	620.12	0.69	49.84	22.79
2018-01-30 09:00:00	145.30	833.22	0.73	35.73	24.12
2018-01-30 10:00:00	149.76	900.95	0.69	21.66	24.87

Dentro de la plataforma Colab, es posible realizar un análisis del set de datos y encontrar características propias de cada variable. Tal y como se describe en la figura 43, todas las variables cuentan con un total de 18,694 de datos, con lo cual no hay una desproporción de cantidad entre estas. El valor medio de potencia eléctrica de la Central Solar Rubí es de un 90.94 MW con un máximo histórico de 157.67 MW, de igual forma la

media de irradiación es de 422.8 Wh/m^2 y contando con un valor máximo de 1112.67 Wh/m^2, representando casi el triple de su valor de media. Esto se debe a que el valor máximo se da en la estación de verano, y se tienen picos menores en lo que resta de cada año, esto genera que su media sea mucho menor y que la desviación estándar sea tan alta. El índice de claridad es de 0.48, esto indica que no es una zona completamente cubierta de nubes, pero no se encuentra ausente de estas, lo cual se ve reflejado en su desviación estándar, manteniendo el valor del índice de claridad dentro de un valor medio, su máximo valor es de 0.83.

De igual forma, el valor del ángulo cenital refleja la posición del sol a lo largo del día, teniendo un valor medio de 47.24, y un máximo de 89.43. La temperatura de la zona de la central solar es de 20.48°C y el valor de la desviación estándar indica que en general esta temperatura se mantiene alrededor de los 20 grados, contando con un máximo de 29°C y un valor mínimo de 10.33°C durante un período de cuatro años. Es importante aclarar que este intervalo de temperaturas se genera a partir de la mitad del día con presencia de la luz solar, no tomando en consideración las temperaturas más bajas que se dan durante las horas de la noche y de madrugada.

El valor mínimo de todas las variables a excepción de la temperatura es cero, esto se debe a que todas estas dependen directamente de la presencia del sol, por lo que este valor mínimo se encuentra durante el amanecer y atardecer del día, esto es alrededor de las horas de inicio y cierre del set de datos que se generó. Asimismo, el valor de desviación estándar y de los percentiles se hubiera visto afectado en el caso en el que se considerara las 24 horas del día, sin embargo, sólo se tomó de actividad solar, donde todas las variables menos la temperatura son iguales a cero.

Figura 43. Análisis del set de datos

	count	mean	std	min	25%	50%	75%	max
Power	18694.0	90.946201	53.727235	0.00	37.2775	115.905	135.2000	157.67
Irradiance	18694.0	422.805754	315.574370	0.00	124.2025	398.670	672.6450	1112.67
Index Clarity	18694.0	0.483441	0.209868	0.00	0.3600	0.500	0.6600	0.83
Zenital Angle	18694.0	47.245574	25.501978	0.00	29.3400	47.590	68.0475	89.43
Temperature	18694.0	20.481540	3.230461	10.33	18.2900	20.650	22.9000	29.06

A partir del uso de las librerías para el manejo de set de datos, se obtiene la correlación de las cinco variables a ingresar al algoritmo, esto se representa en la figura 44. Tal y como se observa, las variables con una mayor correlación directa respecto de la potencia eléctrica son la irradiancia y el índice de claridad, las cuales poseen valores de

0.75 y 0.71 respectivamente. Luego de estas dos variables, se encuentra la temperatura, la cual posee una menor relación positiva directa de 0.55. La variable del ángulo cenital cuenta con un valor mínimo de correlación negativo de -0.27, por lo cual podría considerarse la eliminación de esta variable; sin embargo, en el subsiguiente apartado 5.3.2.1 se observa que, eliminando la variable de temperatura, la tasa de error aumenta en comparación con el modelo entrenado con las cinco variables.

Figura 44. Correlación de parámetros

	Power	Irradiance	Index Clarity	Zenital Angle	Temperature
Power	1.000000	0.754907	0.711749	-0.274669	0.550250
Irradiance	0.754907	1.000000	0.849365	-0.519141	0.735088
Index Clarity	0.711749	0.849365	1.000000	-0.059861	0.663007
Zenital Angle	-0.274669	-0.519141	-0.059861	1.000000	-0.333952
Temperature	0.550250	0.735088	0.663007	-0.333952	1.000000

En las figuras 45, 46, 47, 48 y 49 se muestran las gráficas de las variables con respecto del tiempo, permitiendo observar su comportamiento histórico. En las figuras de potencia e irradiancia se refleja un cambio cíclico, teniendo sus mayores valores a inicios de cada año y los menores a mitades del año.

Figura 45. Histórico de potencia eléctrica

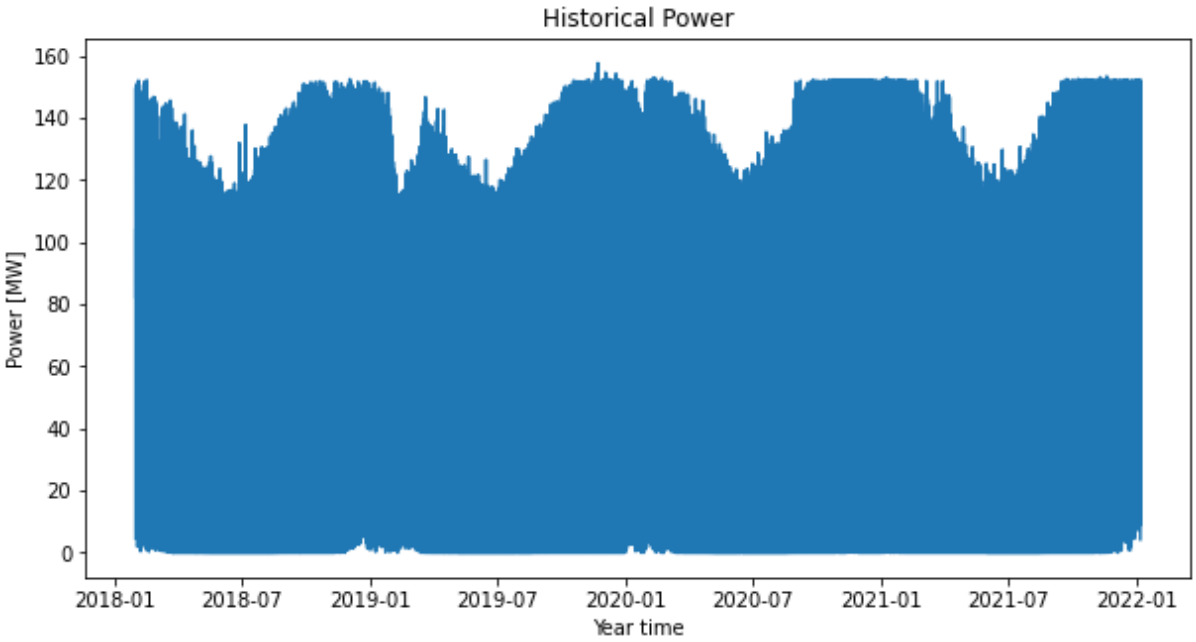
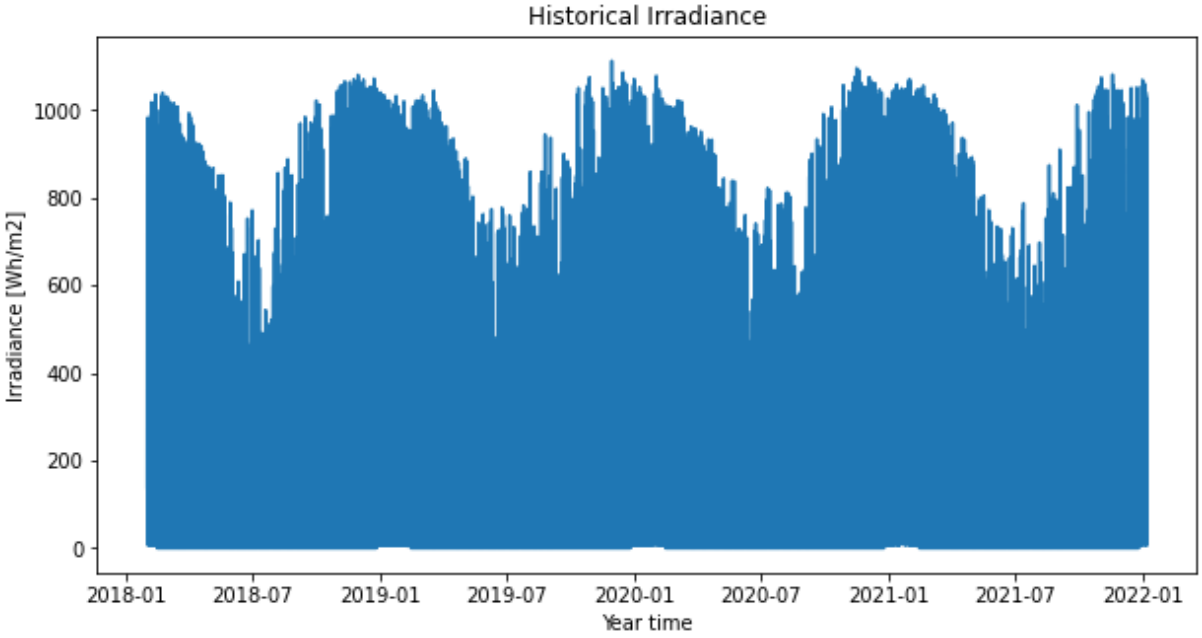


Figura 46. Histórico de irradiancia



En el histórico del índice de claridad no hay una tendencia muy marcada, teniendo muchas variaciones a lo largo del período recopilado. Lo más característico es que el valor máximo oscila entre el 0.8 para luego reducirse a mitades de año por períodos muy esporádicos. A diferencia del índice de claridad, el ángulo cenital tiene una tendencia muy marcada y definida a lo largo de los cuatro años.

Figura 47. Histórico de índice de claridad

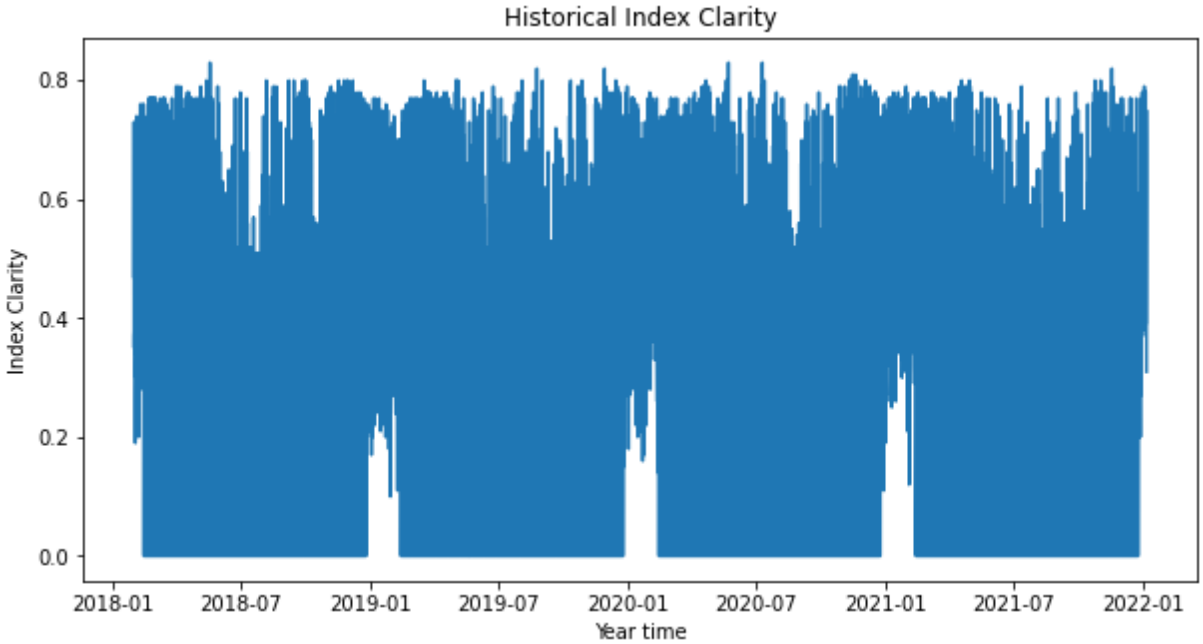
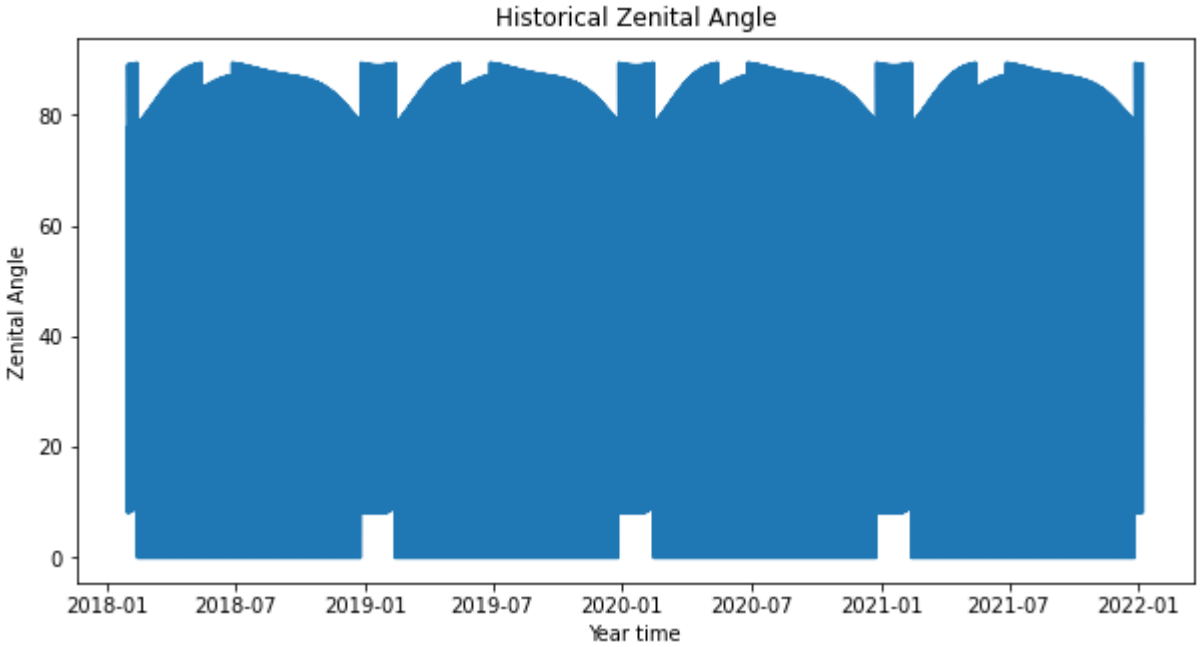
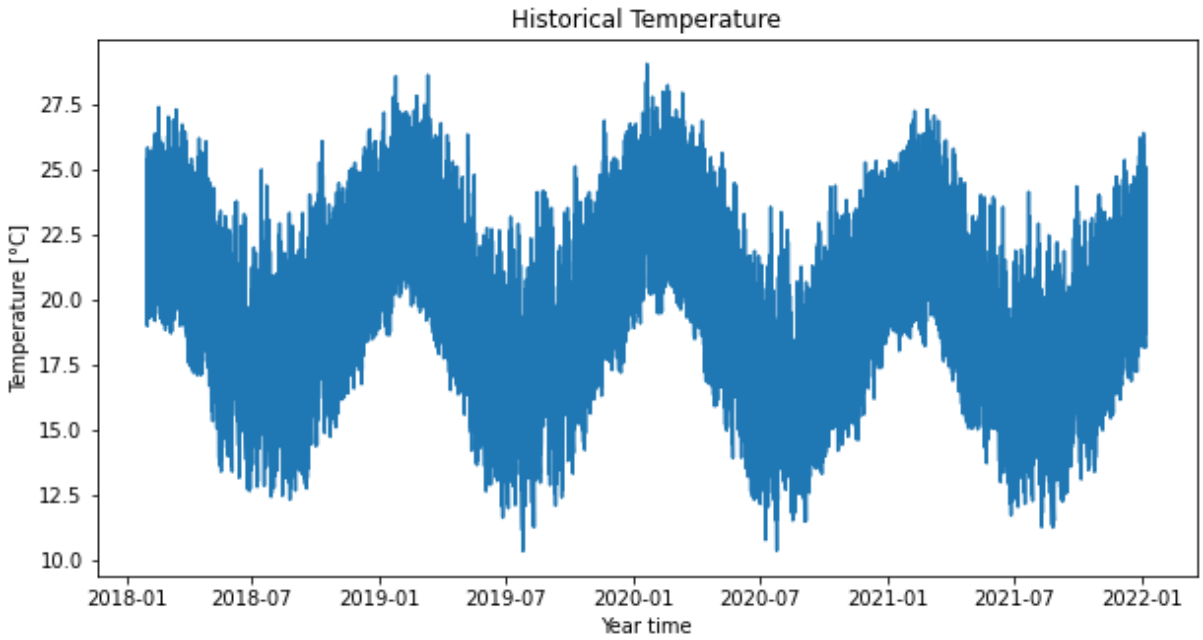


Figura 48. Histórico de ángulo cenital



Finalmente, el histórico de la temperatura durante las horas de generación eléctrica muestra una tendencia cíclica durante el período de los cuatro años de los datos recopilados, esta tendencia se alinea junto con las de potencia e irradiancia.

Figura 49. Histórico de temperatura



Para ingresar esta estructura de datos dentro del algoritmo de predicción LSTM, esta debe ser normalizada, ya que el intervalo de las variables difiere entre sí, el no aplicar este procedimiento al set de datos puede ocasionar una pésima relación entre las propias variables dentro del algoritmo y generar error en los resultados. Por lo que, cada columna de datos es re escalada y sus valores van del cero al uno. Una vez normalizado el set de datos, este queda establecido como se observa en la figura 50.

Figura 50. Set de datos normalizado

Date time	Power	Irradiance	Index Clarity	Zenital Angle	Temperature
2018-01-30 06:00:00	0.029111	0.125167	0.566265	0.869507	0.468767
2018-01-30 07:00:00	0.535612	0.343085	0.746988	0.714302	0.568073
2018-01-30 08:00:00	0.864844	0.557326	0.831325	0.557307	0.665243
2018-01-30 09:00:00	0.921545	0.748847	0.879518	0.399530	0.736252
2018-01-30 10:00:00	0.949832	0.809719	0.831325	0.242201	0.776295

A partir del normalizado del set de datos se realiza la división en sets de entrenamiento, validación y testeo. El set de entrenamiento, como su nombre lo indica, sirve para entrenar a la red neuronal y aprenda de la estructura del set de datos que se ha ingresado, la gran mayoría de los datos recopilados se utiliza en esta parte del algoritmo, en este caso se usó el 80% de todos los datos disponibles. El set de validación es para evaluar el comportamiento del modelo a partir de lo aprendido del set de entrenamiento y; según el resultado, ir modificando el peso de los parámetros de la red para reducir el error de predicción, el 20% de los datos se usa como método de validación. El set de testeo en este estudio es reducido, debido a que el objetivo es poder realizar la predicción de la potencia eléctrica de la central solar Rubí por el período de un día, en este caso la central solar cuenta con alrededor de 12 horas activas de producción.

5.2 Configuración del modelo base

La arquitectura del modelo de predicción LSTM se genera a partir de la librería Tensorflow, especializada en el uso de aprendizaje automático. La capa de entrada del algoritmo sirve como punto de acceso para los datos preparados en etapas previas, en este caso son 6 columnas de variables a ingresar a la capa inicial, la cual es llamada *Input Layer*. Luego de esta capa de entrada, le sigue una capa compuesta por 128 unidades, la cual además tiene activado el retorno de secuencias o *return sequences*, por lo que las salidas de todos los estados ocultos de esta capa serán generadas, lo cual difiere de la

siguiente capa, la cual está compuesta por 64 unidades y el retorno de secuencias está inactivo, dando sólo de retorno la salida final de la secuencia. Ambas capas tienen la función de activación Tanh.

A continuación, se añade el comando *Dropout* con un valor de 0.05, el cual establece aleatoriamente ciertas unidades a un valor de cero para evitar un sobreajuste del modelo. En las últimas dos capas se tienen las funciones de activación Tanh y Linear, de las cuales se obtiene el resultado de la red entrenada, obteniéndose la salida de una sola variable, siendo esta la predicción de potencia eléctrica de la central solar. El modelo LSTM se observa en la figura 51.

Figura 51. Modelo de red LSTM

```
#LSTM
model = Sequential()
model.add(InputLayer((6, 5)))
model.add(LSTM(128, activation='tanh', return_sequences=True))
model.add(LSTM(64, activation='tanh', return_sequences=False))
model.add(Dropout(0.05))
model.add(Dense(8, 'tanh'))
model.add(Dense(1, 'linear'))

model.summary()
```

Para esta arquitectura, se utiliza el optimizador Adamax con un *learning rate* de 0.0001. Los optimizadores son algoritmos o métodos que se encargan de reducir la tasa de error del aprendizaje realizado por el modelo, también llamada función de error. Uno de los más utilizados es el Adamax, por lo cual fue seleccionado para su uso en el modelo inicial. Estas funciones de error dependen directamente de los valores de los pesos y *bias*, pertenecientes a cada unidad del modelo. Este proceso de reducción de error se realiza a partir de un paso, el cual está definido por el hiperparámetro *learning rate*, que busca el mínimo valor para la función de pérdida.

La selección del valor de *learning rate* define la magnitud de cambio que habrá en la búsqueda del valor mínimo para la función error, el utilizar valores altos de *learning rate* ocasiona que no logre converger a un valor mínimo y en el caso de valores bajos puede llevar a la obtención de valores mínimos locales y no conseguir el mínimo general de la función de error. Es debido a esto por lo que se utiliza un valor de 0.0001 para el *learning rate* en la implementación de este modelo.

5.3 Entrenamiento y validación de la red neuronal LSTM con el set de datos

5.3.1 Predicción de potencia eléctrica un día con modelo inicial

Con el modelo base descrito en la sección 5.2, se obtiene los siguientes resultados (figuras 52 y 53):

Figura 52. Reducción de error del modelo inicial

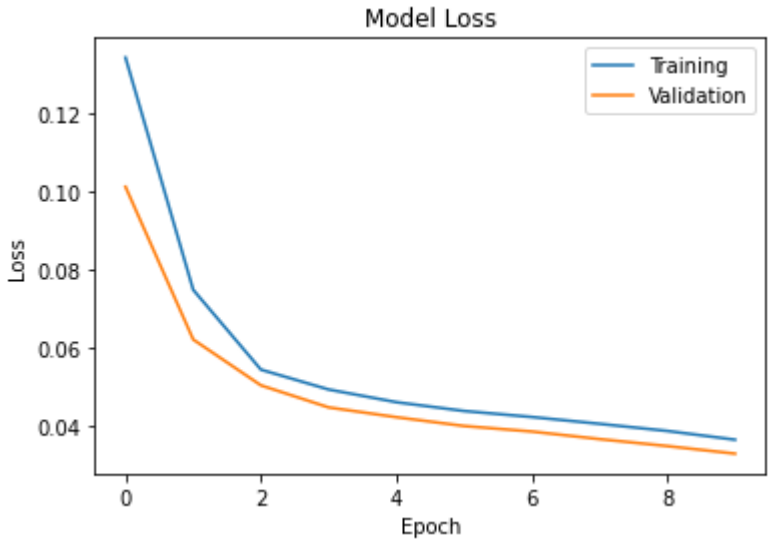
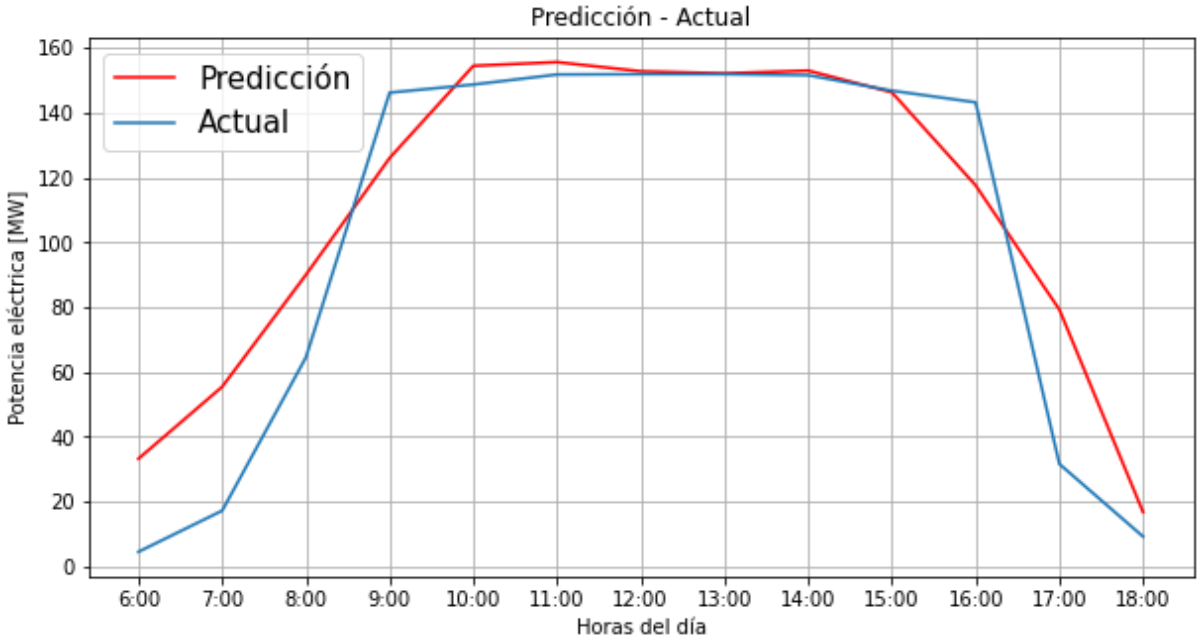


Figura 53. Predicción y valor real de potencia eléctrica del modelo inicial



En la tabla 7 se observa los valores de error alcanzados por el modelo, estos son medidos a partir de las siguientes métricas: MAE (*Mean of Absolute value of Errors*), el cual mide el error promedio entre los valores reales y los valores obtenidos por el modelo de predicción mientras que el RMSE (*Root of the Mean of the Square of Errors*) es la raíz cuadrada de la diferencia cuadrática de la diferencia de los valores reales y los pronosticados.

Tabla 7 Resultados del modelo inicial

	MAE	RMSE
Modelo inicial	0.1443	0.1813

Sumado a estas métricas, se utilizarán otros factores con el fin de medir la performance del modelo, uno de estos es el error del valor de energía eléctrica generada por la central solar, a partir de la comparativa del modelo frente al valor real. Adicional a esto, se considera el porcentaje de las pendientes al inicio del día, entre las seis y ocho de la mañana, y al final del día, entre las cuatro y seis de la tarde.

El porcentaje de error de la energía conseguido por el modelo sirve como indicativo para medir el nivel de precisión en la sumatoria de la producción diaria de la central solar en comparativa al valor real. Sin embargo, el modelo puede tener dificultades en seguir la tasa de cambio en horas donde el aumento o disminución son bastante marcadas, por lo que el porcentaje de error de las pendientes muestra la exactitud del modelo en seguir los cambios.

En la tabla 8 se muestran los resultados referente a los valores de energía y pendientes del modelo inicial. Hay un error porcentual de 8.61% referente a la energía, mientras que la tasa de error se observa que en la pendiente inicial hay un error del 5.41% y de la pendiente final es de 24.75%. Si bien el modelo presenta pendientes similares al inicio de la predicción, el valor inicial de potencia eléctrica difiere del valor real, siendo la predicción de 33 MW y el valor real 4.5 MW.

Tabla 8. Resultados de energía y pendientes modelo inicial

Modelo inicial	Predicción	Real	Error (%)
Energía (MWh)	1432.61	1319.1	8.61
Pendiente inicial	28.37	30	5.41
Pendiente final	-50.41	-67	24.75

5.3.2 Predicción de potencia eléctrica un día con modelo optimizado

El modelo inicial da unos buenos resultados y logra predecir la generación de potencia eléctrica de la central solar Rubí; sin embargo, la tasa de error del modelo puede ser mejorada y generar una predicción más precisa, por lo que algunos de los hiperparámetros del modelo son modificados. En vez de utilizar la función de activación Tanh como se hizo inicialmente, se utilizó otras funciones para conseguir la que mejor se adapte al modelo, la función ReLu fue la que dio mejores resultados. A su vez, el comando Dropout es eliminado de la arquitectura del modelo, esto se realiza con el objetivo de darle un peso y *bias* a todas las unidades de la red neuronal y eliminar el porcentaje de unidades con un valor de cero que sin importar los cambios en las funciones de error mantenían este valor, esto genera que las características de las variables influyan de manera general en la red neuronal. En el caso de este modelo, la eliminación del comando Dropout mejora los resultados.

De igual forma, se utilizó otro optimizador, llamado Adam, el cual consigue mejores valores en la reducción de la función de error en comparación del optimizador Adamax, en base a esto se realiza el cambio de optimizador. Se utilizaron distintos valores de *learning rate*; sin embargo, el valor inicial utilizado es el que genera mejores resultados, por lo que se mantiene este valor. A partir de los cambios mencionados se obtuvieron los siguientes resultados representados en las figuras 54 y 55:

Figura 54. Reducción de error del modelo optimizado

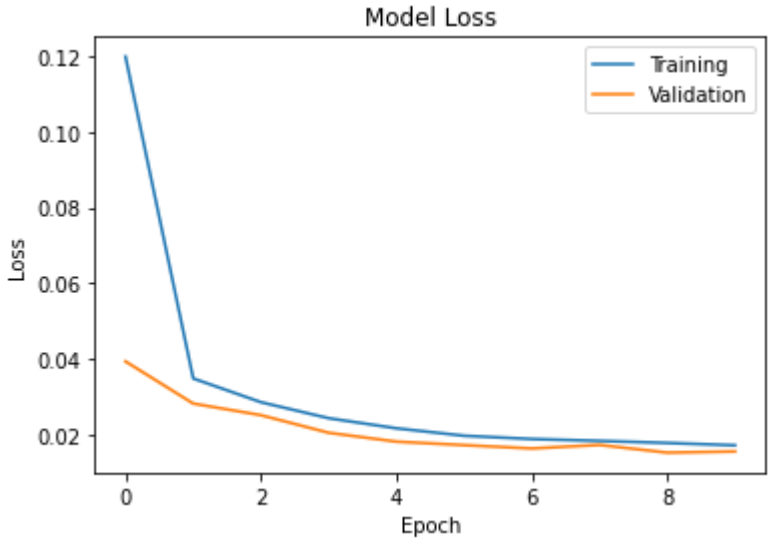
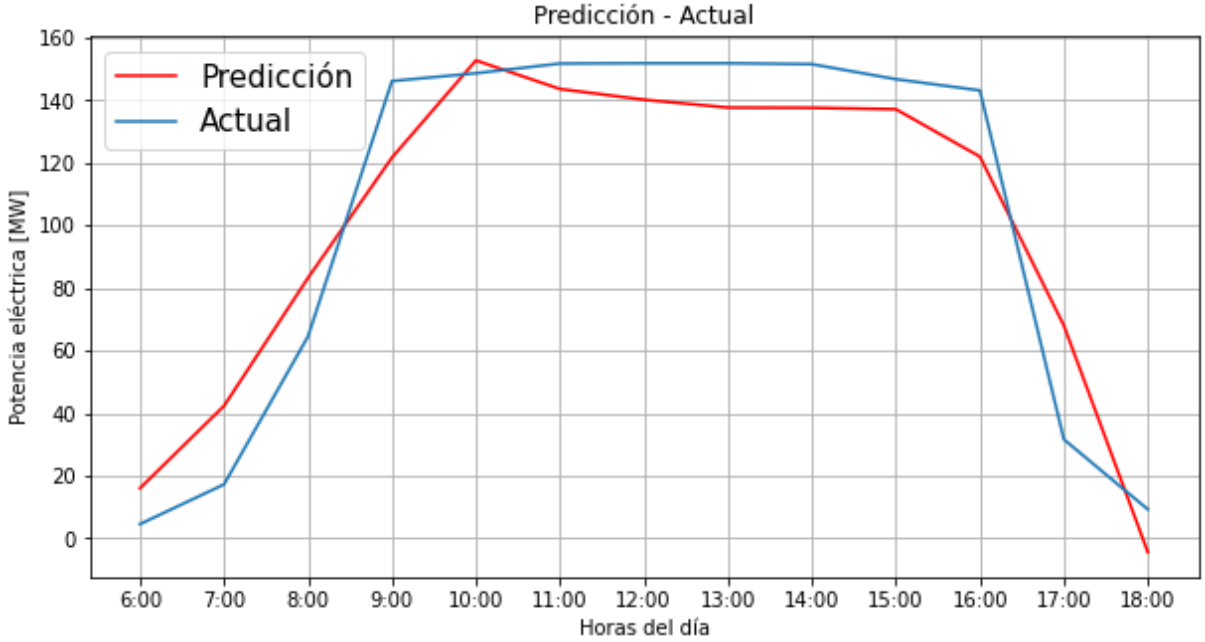


Figura 55. Predicción y valor real de potencia eléctrica del modelo optimizado



En la tabla 9 se observan los valores de error alcanzados por el modelo optimizado, este cuenta con un valor de MAE de 0.0829 y de RMSE tiene un resultado de 0.122, lo cual representa una mejora en la tasa de error en comparación del modelo inicial.

Tabla 9 Resultados del modelo optimizado

	MAE	RMSE
Modelo optimizado	0.0829	0.122

En la tabla 10 se muestran los resultados referente a los valores de energía y pendientes del modelo inicial. Hay un error porcentual de 1.59% referente a la energía, mientras que la tasa de error se observa que en la pendiente inicial hay un error del 12.06% y de la pendiente final es de 5.55%. El modelo optimizado presenta de igual manera dificultad en predecir el valor inicial de potencia eléctrica con el cual empieza el día, al igual que en las horas pico, el valor de la predicción es inferior al alcanzado por la central solar.

Tabla 10. Resultados de energía y pendientes modelo optimizado

Modelo inicial	Predicción	Real	Error (%)
Energía (MWh)	1298.1	1319.1	1.59
Pendiente inicial	33.61	30	12.06
Pendiente final	-63.28	-67	5.55

5.3.2.1 Modelo optimizado utilizando cuatro variables. A partir del cuadro de correlación de las variables, se observa que hay un bajo valor de correlación del ángulo cenital con respecto de la potencia eléctrica de la central. Por lo que, para medir el impacto que esta variable tiene sobre el modelo, en este apartado el modelo predictivo tendrá sólo cuatro variables de input, eliminando el ángulo cenital. Los resultados se observan en las figuras 56 y 57.

Figura 56. Reducción de error del modelo optimizado con cuatro variables

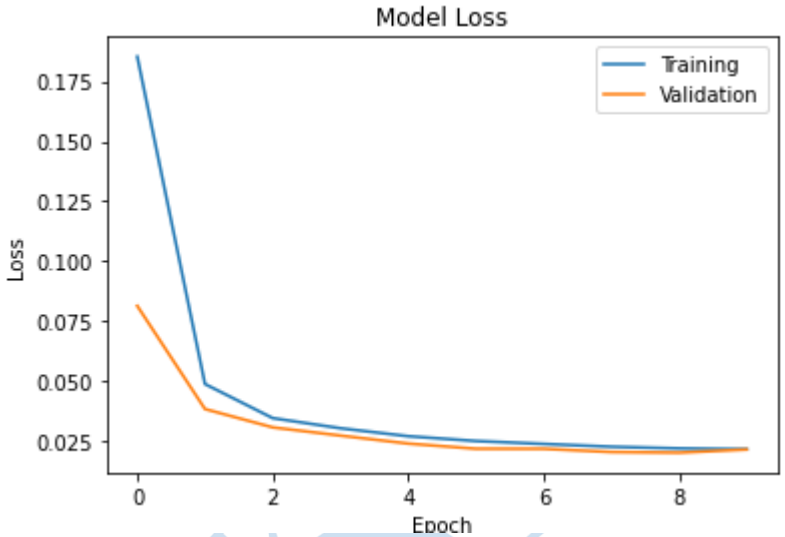
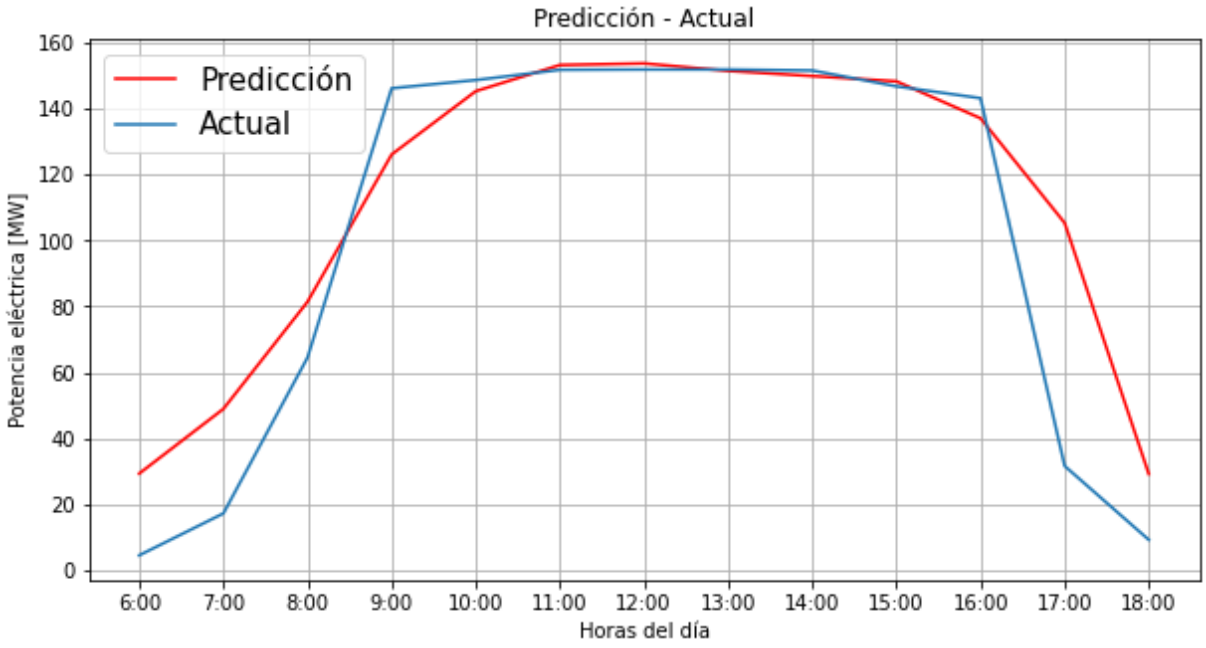


Figura 57. Predicción y valor real de potencia eléctrica del modelo optimizado con cuatro variables



En la tabla 11 se observan los valores de error alcanzados por el modelo optimizado usando sólo cuatro variables, el resultado obtenido da unos valores MAE de 0.1136 y RMSE de 0.1464, lo cual representa un aumento en la tasa de error en comparación del modelo entrenado con las cinco variables.

Tabla 11. Resultados del modelo optimizado con cuatro variables

	MAE	RMSE
Modelo optimizado con cuatro variables	0.1136	0.1464

En la tabla 12 se muestran los resultados referente a los valores de energía y pendientes del modelo inicial. Hay un error porcentual de 5.52% referente a la energía, mientras que la tasa de error se observa que en la pendiente inicial hay un error del 4.9% y de la pendiente final es de 10.8%. El modelo con input de las cuatro variables elegidas tiene un mayor error al predecir la energía total y el porcentaje de error de las pendientes de inicio y fin del día son bajas, sin embargo, no logra llegar y mantenerse en el valor de producción máximo que tiene la central en su mayoría de horas.

Tabla 12. Resultados de energía y pendientes del modelo optimizado con cuatro variables

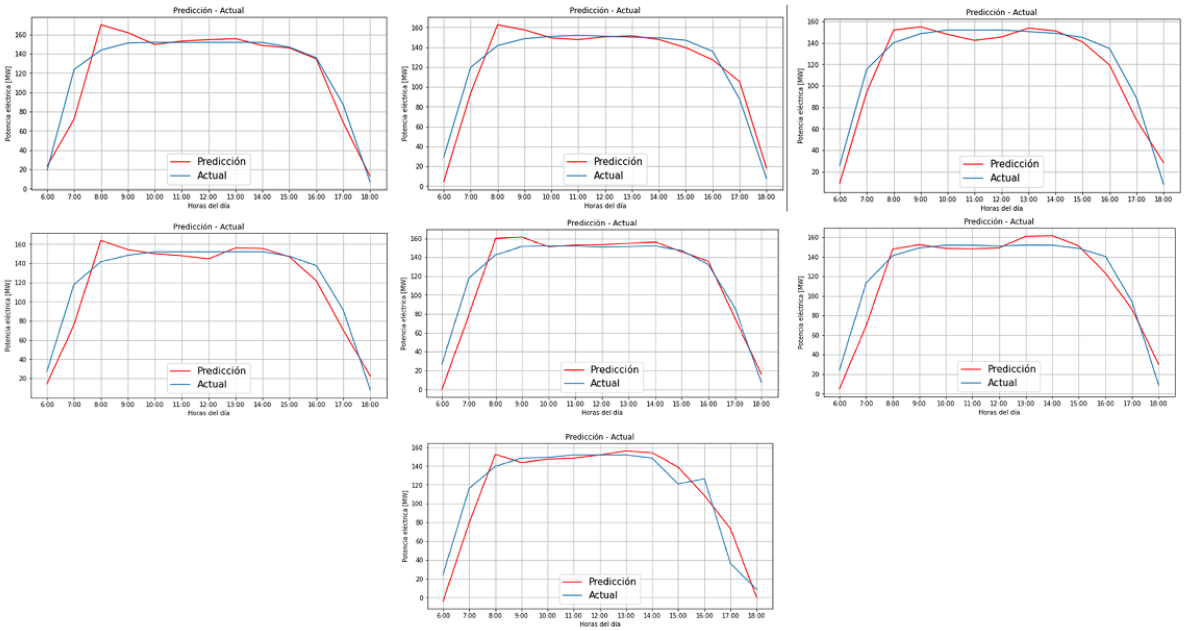
Modelo optimizado con cuatro variables	Predicción	Real	Error (%)
Energía (MWh)	1459.43	1319.1	10.64
Pendiente inicial	26.1	30	13.13
Pendiente final	-53.97	-67	19.44

5.3.3 Predicción de potencia eléctrica de siete días

A partir del modelo optimizado generado en el apartado 5.3.2, se procederá a realizar la predicción de potencia eléctrica de una semana. Este proceso se llevará a cabo a partir de la predicción de generación de un día, luego, los datos recopilados de este día serán ingresados al set de validación del modelo parar generar la predicción del día siguiente, esta ejecución será realizada hasta llegar al séptimo día. Con este proceso de predicción se busca ver si al añadir datos de cortos períodos de tiempo influyen en la performance del modelo de predicción LSTM.

Los resultados de predicción junto con los valores reales de potencia se detallan en las figuras 58, representado por el valor real y la predicción del modelo en los siete días.

Figura 58. Predicción y valor real de potencia eléctrica de una semana



5.4 Resultados

En la tabla 13 y 14 se detalla la tasa de error obtenida tanto para los modelos desarrollados en el apartado 5.3. Se aprecia una notable reducción en la tasa de error entre el modelo inicial y final, llegando a reducirse el MAE de un valor de 0.1443 a 0.0829 y el RMSE de 0.1813 a 0.122 respectivamente. Esta mejora del desempeño del modelo permite comprender la importancia de establecer correctamente los hiperparámetros de la arquitectura del algoritmo, seleccionando una función de activación y optimizador que logre entregar predicciones con menor tasa de error.

En base al proceso realizado en el apartado 5.3.2 para los siete días de predicción, no puede ser atribuida una mejora directa a la tasa de error a partir de la adición de los datos de un solo día, ya que el error incrementa o disminuye conforme se van añadiendo los datos de los días respectivos al set de validación. Los valores de RMSE tiene un valor máximo de 0.1409 y un valor mínimo de 0.1239. De igual forma en la predicción de energía o en el porcentaje de error de las pendientes no hay una visible mejora conforme se van añadiendo nuevos días al set de entrenamiento. Este resultado hace inferir que, para obtener una mejora considerable del modelo, es necesario añadir una cantidad considerable de datos, en adición a la existente.

Tabla 13 Resultados de MAE y RMSE

Predicción de un día	MAE	RMSE
Modelo inicial	0.1443	0.1813
Modelo optimizado	0.0829	0.122
Modelo optimizado con cuatro variables	0.1136	0.1464
Predicción de siete días	MAE	RMSE
Primer día	0.085	0.1227
Segundo día	0.0865	0.1265
Tercer día	0.0931	0.1302
Cuarto día	0.0962	0.1295
Quinto día	0.0967	0.1320
Sexto día	0.0997	0.1351
Séptimo día	0.0867	0.1249

Tabla 14. Resultados de energía y pendientes

Predicción de un día			
Modelo inicial	Predicción	Real	Error (%)
Energía (MWh)	1194.66	1319.1	9.38
Pendiente inicial	16.49	18.6	11.65
Pendiente final	-22.04	-40	44.89

Predicción de un día			
Modelo optimizado	Predicción	Real	Error (%)
Energía (MWh)	1302.3	1319.1	1.27
Pendiente inicial	37.71	47.23	20.14
Pendiente final	-44.394	-45.86	3.21
Modelo optimizado con cuatro variables	Predicción	Real	Error (%)
Energía (MWh)	1194.66	1319.1	9.38
Pendiente inicial	16.49	18.6	11.65
Pendiente final	-22.04	-40	44.89
Predicción de siete días			
Primer día	Predicción	Real	Error (%)
Energía (MWh)	1553.23	1575.4	1.41
Pendiente inicial	73.23	62.25	17.64
Pendiente final	-60.88	-64.4	5.46
Segundo día	Predicción	Real	Error (%)
Energía (MWh)	1555.07	1570.8	1.01
Pendiente inicial	79.07	56.25	40.58

Predicción de siete días			
Segundo día	Predicción	Real	Error (%)
Pendiente final	-54.60	-64.05	14.74
Tercer día	Predicción	Real	Error (%)
Energía (MWh)	1508.03	1562.5	3.48
Pendiente inicial	71.31	57.05	25.01
Pendiente final	-45.4	-63.2	28.16
Cuarto día	Predicción	Real	Error (%)
Energía (MWh)	1522.67	1578.5	3.54
Pendiente inicial	74.69	57.05	30.93
Pendiente final	-49.61	-64.89	23.56
Quinto día	Predicción	Real	Error (%)
Energía (MWh)	1541.03	1569.2	1.79
Pendiente inicial	79.93	57.9	38.05
Pendiente final	-59.65	-62.4	4.39

Predicción de siete días			
Sexto día	Predicción	Real	Error (%)
Energía (MWh)	1533.34	1578	2.83
Pendiente inicial	71.38	58.49	22.01
Pendiente final	46.6	65.8	29.12
Séptimo día	Predicción	Real	Error (%)
Energía (MWh)	1450.35	1474	1.60
Pendiente inicial	78.15	57.60	35.68
Pendiente final	54.27	58.9	7.86

UNIVERSITAS STUDIORUM PIURENSIS

Conclusiones

El uso de tecnologías de predicción con base en algoritmos de inteligencia artificial acelera la inclusión de generadoras de energía renovables a las redes nacionales, permitiendo así tener un pronóstico de producción eléctrica en el horizonte de tiempo solicitado por la coordinadora nacional, reduciendo la incertidumbre de su participación programada.

Una empresa de generación eléctrica puede verse beneficiada económicamente si esta cuenta con un correcto sistema de predicción, el cual le permitirá fijar un precio competitivo en el mercado a partir de la energía a producir. Los sistemas de predicción cobran mayor relevancia para las generadoras de energía renovables debido a la naturaleza volátil que estas presentan.

La comparativa entre la predicción de la potencia eléctrica del modelo junto con la generación real de la central, puede utilizarse como un indicativo o alarma que algunos paneles no están funcionando según lo esperado, por lo que debería realizarse un mantenimiento correctivo.

El manejo de datos a entrenar es el punto clave para el desarrollo de un correcto modelo de predicción, ya que a partir de este proceso se recopila, analiza y limpia el set de datos, eliminando posibles errores que generen resultados de predicción con un margen de error elevado.

La performance del modelo de predicción basado en redes neuronales LSTM se ve influida en gran parte por la cantidad de datos que se dispongan. Esto es debido a que el algoritmo LSTM mantiene lo aprendido en el largo plazo, a diferencia de una red recurrente normal, por lo que, para visualizar mejora o cambios en los resultados de predicción, la cantidad de datos a añadir debe ser considerable. Inicialmente estos datos nuevos formarán parte del aprendizaje a corto plazo, los cuales progresivamente se adicionarán al aprendizaje a largo plazo.

UNIVERSITAS STUDIORUM PIURENSIS

Recomendaciones

Utilizar los datos recopilados por la propia central solar Rubí, ya que estos resultarán más útiles y precisos. La información recopilada por parte de la central permitirá obtener una mayor cantidad de variables a usar para el modelo de predicción y a su vez, reducir el intervalo de tiempo que hay entre cada toma de datos. Una de las variables a adicionar sería la temperatura propia de los paneles, esta variable en comparativa a la temperatura ambiental podría tener una correlación positiva directa respecto a la potencia eléctrica de la central solar.

Generar un proceso automatizado de adición de datos, permitiendo actualizar la base de datos de la central de manera diaria, y a partir de la recopilación de datos por un período de tiempo considerable, actualizar el modelo de red neuronal con el fin de mejorar los resultados de predicción.

Aplicar este modelo de predicción de potencia eléctrica LSTM a otras centrales solares, u otras centrales de generación renovable, como las centrales eólicas o hidráulicas. La selección de las variables a utilizar para la predicción dependerá del tipo de central.

UNIVERSITAS STUDIORUM PIURENSIS

Referencias bibliográficas

Aghaei, M., Eskandari, A., Vaezi, S., & Chopra, S. S. (2020). Solar PV power plants. In *Photovoltaic Solar Energy Conversion* (pp. 313–348). Elsevier. https://doi.org/10.1016/B978-0-12-819610-6.00010-7

Auffarth, B. (2021). *Machine Learning for Time-Series with Python*.

Bonan, G. (2015). Ecological Climatology. In *Our Warming Planet: Topics In Climate Dynamics*. Cambridge University Press. https://doi.org/10.1017/CBO9781107339200

Chollet, F. (2018). Deep Learning with Python. In *MANNING SHELTER ISLAND*.

Cielen, D., Meysman, A. D. B., & Ali, M. (2016). *Introducing Data Science*. Manning Publications Co.

COES. (2022a). *Informe de Evaluación de la Operación Diaria*. https://www.coes.org.pe/Portal/PostOperacion/Reportes/Ieod

COES. (2022b). *Informe De La Operacion Anual Del Sein 2021*.

COES. (2022c). *Organización*. https://www.coes.org.pe/Portal/Organizacion/QuienesSomos

Enel. (2018). *Rubí, energía solar para una nueva era*. https://www.enel.pe/es/sostenibilidad/rubi-la-central-solar-de-enel-mas-grande-del-peru.html

ESMAP. (2020). *Global Photovoltaic Power Potential by Country*. https://globalsolaratlas.info/global-pv-potential-study

European Comimission. (2019). Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission. A definition of AI: Main capabilities and scientific disciplines High-Level. *European Commission*, 7. https://ec.europa.eu/digital-single-

Fouad, M. M., Shihata, L. A., & Morgan, E. S. I. (2017). An integrated review of factors influencing the performance of photovoltaic panels. *Renewable and Sustainable Energy Reviews*, *80*(June), 1499–1511. https://doi.org/10.1016/j.rser.2017.05.141

Géron, A. (2011). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow 2nd Edition (2019). In *Journal of Physics A: Mathematical and Theoretical* (Vol. 44, Issue 8).

Green, M. A., Dunlop, E. D., Hohl-Ebinger, J., Yoshita, M., Kopidakis, N., & Ho-Baillie, A. W.

Y. (2020). Solar cell efficiency tables (Version 55). *Progress in Photovoltaics: Research and Applications*, *28*(1), 3–15. https://doi.org/10.1002/pip.3228

IRENA. (2020). Renewable Power Generation Costs in 2019. In *Renewable Power Generation Costs in 2019*. International Renewable Energy Agency (IRENA). https://www.irena.org/publications/2020/Jun/Renewable-Power-Costs-in-2019

Kelleher, J. D., & Tierney, B. (2018). DATA SCIENCE. In *Massachusetts Institute of Technology All*. https://lccn.loc.gov/2017043665

Liu, C.-H., Gu, J.-C., & Yang, M.-T. (2021). A Simplified LSTM Neural Networks for One Day-Ahead Solar Power Forecasting. *IEEE Access*, *9*, 17174–17195. https://doi.org/10.1109/ACCESS.2021.3053638

Lynn, P. A. (2010). Electricity from Sunlight. In *Electricity from Sunlight*. https://doi.org/10.1002/9780470710111

Martín, A. M., & Dominguez, J. (2019). Solar Resources Mapping. In J. Polo, L. Martín-Pomares, & A. Sanfilippo (Eds.), *Green Energy and Technology*. Springer International Publishing. https://doi.org/10.1007/978-3-319-97484-2

Mishra, M., Byomakesha Dash, P., Nayak, J., Naik, B., & Kumar Swain, S. (2020). Deep learning and wavelet transform integrated approach for short-term solar PV power prediction. *Measurement: Journal of the International Measurement Confederation*, *166*, 108250. https://doi.org/10.1016/j.measurement.2020.108250

Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. In *Machine Learning: A Probabilistic Perspective* (p. 1098). Massachusetts Institute of Technology. https://mitpress.mit.edu/9780262304320/machine-learning/

NASA. (2022). *What Is NASA?*

OSINERGMIN. (2019). *Osinergmin Energias Renovables Experiencia Perspectivas*. 143. https://www.gob.pe/institucion/osinergmin/informes-publicaciones/483400-energias-renovables-experiencia-y-perspectivas-en-la-ruta-del-peru-hacia-la-transicion-energetica

OSINERGMIN. (2021). *SUPERVISIÓN DE CONTRATOS DE PROYECTOS DE GENERACIÓN Y TRANSMISIÓN DE ENERGÍA ELÉCTRICA EN OPERACIÓN*.

Perpiñán, O. (2013). *ENERGÍA SOLAR*.

Qing, X., & Niu, Y. (2018). Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy*, *148*, 461–468. https://doi.org/10.1016/j.energy.2018.01.177

Rodríguez, F., Martín, F., Fontán, L., & Galarza, A. (2021). Ensemble of machine learning and spatiotemporal parameters to forecast very short-term solar irradiation to compute photovoltaic generators' output power. *Energy*, *229*. https://doi.org/10.1016/j.energy.2021.120647

Satpathy, R., & Pamuru, V. (2021). Grid integration, performance, and maintenance of solar PV power systems. In *Solar PV Power* (pp. 435–481). Elsevier. https://doi.org/10.1016/B978-0-12-817626-9.00010-1

Sharadga, H., Hajimirza, S., & Balog, R. S. (2020). Time series forecasting of solar power generation for large-scale photovoltaic plants. *Renewable Energy*, *150*, 797–807. https://doi.org/10.1016/j.renene.2019.12.131

Stuart, R., & Peter, N. (2021). *Artificial Intelligence: A Modern Approach* (Issue 1). www.pearsonglobaleditions.com

The POWER Project. (2022).

United Nations Statistics División. (2021). the Energy Progress Report 2021. *Iea*. https://trackingsdg7.esmap.org/data/files/download-documents/2021_tracking_sdg7_chapter_6_outlook_for_sdg7.pdf

Wang, K., Qi, X., & Liu, H. (2019). Photovoltaic power forecasting based LSTM-Convolutional Network. *Energy*, *189*, 116225. https://doi.org/10.1016/j.energy.2019.116225

Watt, J., Borhani, R., & Katsaggelos, A. (2020). Machine Learning Refined. In *Cambridge University Press* (Vol. 91). Cambridge University Press. https://doi.org/10.1017/9781108690935

Yin, R., Li, D., Wang, Y., & Chen, W. (2020). Forecasting method of monthly wind power generation based on climate model and long short-term memory neural network. *Global Energy Interconnection*, *3*(6), 571–576. https://doi.org/10.1016/j.gloei.2021.01.003

Yu, M., Xu, F., Hu, W., Sun, J., & Cervone, G. (2021). Using Long Short-Term Memory (LSTM) and Internet of Things (IoT) for localized surface temperature forecasting in an urban environment. *IEEE Access*, *9*, 137406–137418. https://doi.org/10.1109/ACCESS.2021.3116809

Zekry, A., Shaker, A., & Salem, M. (2018). Solar Cells and Arrays: Principles, Analysis, and Design. In *Advances in Renewable Energies and Power Technologies* (Vol. 1). Elsevier Inc. https://doi.org/10.1016/B978-0-12-812959-3.00001-0

UNIVERSITAS STUDIORUM PIURENSIS

Apéndices

UNIVERSITAS STUDIORUM PIURENSIS

Apéndice A. Código LSTM

```
import tensorflow as tf

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/Central Solar Rubi.csv")

df

df.dtypes

df['Date time'] = pd.to_datetime(df['Date time'], errors='coerce')

del df["Date"]

del df["Year"]

del df["Month"]

del df["Day"]

del df["Hour"]

df.describe().transpose()

df.corr()

df.head()

power = df['Temperature']

power.index = df['Date time']

plt.figure(figsize=(10, 5))

plt.title('Historical Temperature')

plt.xlabel('Year time')

plt.ylabel('Temperature')

display(plt.plot(power))

df['Zenital Angle'].plot(figsize=(12, 4), subplots=True) [:5]

plt.show()

plot_cols1 = ['Power', 'Irradiance', 'Index Clarity', 'Zenital Angle', 'Temperature']

new = df[plot_cols1]

new.index = df['Date time']

new

from sklearn.metrics import mean_squared_error as mse

def plot_predictions1(model, X, y, start=0, end=25):

  predictions = model.predict(X).flatten()

  df = pd.DataFrame(data={'Predictions':predictions, 'Actuals':y})
```

```
  plt.grid()
  plt.plot(df['Predictions'][start:end])
  plt.plot(df['Actuals'][start:end])
  return df, mse(y, predictions)
def df_to_X_Y(df, window_size=6):
  df_as_np = df.to_numpy()
  X = []
  Y = []
  for i in range(len(df_as_np)-window_size):
    row = [r for r in df_as_np[i:i+window_size]]
    X.append(row)
    label = df_as_np[i+window_size][0]
    Y.append(label)
  return np.array(X), np.array(Y)
new.min()
new.max()
new = (new-new.min(axis=0))/ (new.max(axis=0) - new.min(axis=0))
new
X, Y = df_to_X_Y(new)
X.shape, Y.shape
endval_begtest=18662
endtest=endtest=18675
X_train, Y_train = X[:15000], Y[:15000]
X_val, Y_val = X[15000:endval_begtest], Y[15000:endval_begtest]
X_test, Y_test = X[endval_begtest:endtest], Y [endval_begtest:endtest]
X_train.shape,  Y_train.shape,  X_val.shape,  Y_val.shape,  X_test.shape,
Y_test.shape
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import *
from tensorflow.keras.callbacks import ModelCheckpoint
from tensorflow.keras.losses import MeanSquaredError
from tensorflow.keras.metrics import RootMeanSquaredError
from tensorflow.keras.optimizers import Adam
#LSTM
model = Sequential()
model.add(InputLayer((6, 5)))
```

```
model.add(LSTM(128, activation='relu', return_sequences=True))
model.add(LSTM(64, activation='relu', return_sequences=False))
#model.add(Dropout(0.05))
model.add(Dense(8, 'relu'))
model.add(Dense(1, 'linear'))
model.summary()
cp = ModelCheckpoint('cp/', save_best_only=True)
model.compile(loss=MeanSquaredError(),optimizer=Adam(learning_rate=0.0001),
metrics=[tf.keras.metrics.MeanAbsoluteError(),tf.keras.metrics.RootMeanSqua
redError(name='root_mean_squared_error', dtype=None)])
hist = model.fit(X_train, Y_train, validation_data=(X_val, Y_val), epochs=10,
callbacks=[cp])
plt.figure()
plt.plot(hist.history['loss'])
plt.plot(hist.history['val_loss'])
plt.title('Model Loss')
plt.ylabel('Loss')
plt.xlabel('Epoch')
plt.legend(['Training','Validation'], loc = 'upper right')
plt.show()
plot_predictions1(model, X_test, Y_test)
salida =[]
def plot_predictions2(model, X, y, start=0, end=25):
  predictions = model.predict(X).flatten()
  Prediccion    =    predictions*(new2['Power'].max()-new2['Power'].min())+
new2['Power'].min()
  Actuales=y*(new2['Power'].max()-new2['Power'].min())+ new2['Power'].min()
  df = pd.DataFrame(data={'Predictions':Prediccion, 'Actuals':Actuales})
  salida.append(df)
  plt.grid()
  plt.title('Predicción-Actual')
  plt.ylabel('Potencia eléctrica [MW]')
  plt.xlabel('Horas del día')
  plt.legend(['Predictions','Actuals'], loc = 'upper left')
  plt.plot(df['Predictions'][start:end],"r")
  plt.plot(df['Actuals'][start:end])
  return df, mse(y, predictions)
```

```
salida
ED=np.squeeze(salida, axis=0)
ED
Valores=pd.DataFrame(ED, columns=['Predictions','Actuals'])
Valores
Valores.dtypes
Final=pd.concat([Valores, Hours],axis=1)
Final.index = Final['Hour']
Final
Final.sum()
def sumatoria():
  suma=0
  sumaR=0
  for i in range(13):
    suma+=Final.iloc[i,0]
  for i in range(13):
    sumaR+=Final.iloc[i,1]
  porcentaje=((sumaR-suma)/sumaR)*100
  return suma,sumaR,porcentaje
sumatoria()
plt.figure(figsize=(10,5))
plt.title('Predicción - Actual')
plt.ylabel('Potencia eléctrica [MW]')
plt.xlabel('Horas del día')
plt.grid()
plt.plot(Final['Predictions'][0:13],"r", label="Predicción")
plt.plot(Final['Actuals'][0:13], label="Actual")
plt.legend(prop={'size':15})
plt.show()
def pen():
  Pen_Predictiva_I= (Final.iloc[2,0]-Final.iloc[0,0])/2
  Pen_Real_I= (Final.iloc[2,1]-Final.iloc[0,1])/2
  Pen_Predictiva_F= (Final.iloc[12,0]-Final.iloc[10,0])/2
  Pen_Real_F= (Final.iloc[12,1]-Final.iloc[10,1])/2
  porcentaje_I= ((Pen_Real_I-Pen_Predictiva_I)/Pen_Real_I)*100
  porcentaje_F=((Pen_Real_F-Pen_Predictiva_F)/Pen_Real_F )*100
```

```
  return   Pen_Predictiva_I,   Pen_Real_I,   Pen_Predictiva_F,   Pen_Real_F,
porcentaje_I, porcentaje_F

pen()
```

UNIVERSITAS STUDIORUM PIURENSIS

Anexos

UNIVERSITAS STUDIORUM PIURENSIS

Anexo A. Hoja informativa OSINERGMIN

Osinergmin
Organismo Supervisor de la Inversión en Energía y Minería

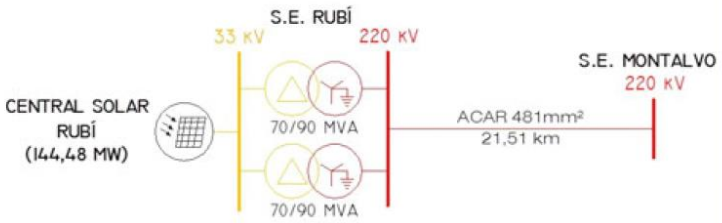
CENTRAL SOLAR RUBÍ (144,48 MW)

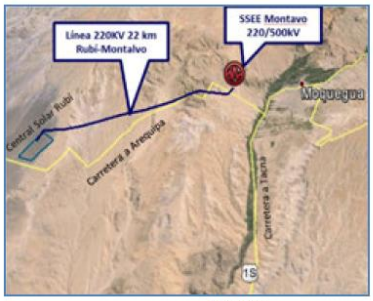
EMPRESA CONCESIONARIA	ENEL GREEN POWER PERÚ S.A.		
DESCRIPCIÓN			
La central tendrá una capacidad de 144,48 MW, que se obtendrá mediante la instalación de 560880 módulos fotovoltaicos de 320 W. La conexión al SEIN se realizará a través dela L.T. 220 kV S.E. Rubí - S.E. Montalvo, de simple circuito de 21,51 km.			
UBICACIÓN			
Departamento	Moquegua		
Provincia	Mariscal Nieto		
Distrito	Moquegua		
Altitud	1 410 msnm		
DATOS DE LA CENTRAL			
Potencia Instalada	144,48 MW		
N° de Módulos en serie	30		
Tipo de Módulos	Reisen		
Potencia de Módulos	320 W		
N° de Inversores	164		
Tipo de Inversores	Fimer R11015TL		
Potencia de Inversores	1,025 MVA		
Tensión de Salida de Inversores	0,4 kV - AC		
Tipo de Estructura	Seguidor Horizontal de 1 eje		
Ángulo de Seguimiento	45°		
N° de Centros de Transformación	41		
Potencia por Centro de Transfor.	3,524 MW		
Relación de Transformación	0,4/33 KV		
DATOS DEL TRANSFORMADOR	**T1**	**T2**	
Potencia Nominal	70/90 MVA	70/90 MVA	
Relación de Transformación	33/220 kV	33/220 kV	
Marca	Chint Eléctric	Chint Eléctric	
Año de Fabricación	2017	2017	
DATOS DE CONTRATO		**HITOS**	
Tipo de Contrato	Suministro RER (4ta Subasta)	Cierre Financiero	15.12.2015 (si)
Firma de Contrato	17.05.2016	Llegada de Equipos	18.08.2017 (si)
Energía Ofertada	415,00 GWh/año	Inicio de Obras	30.11.2016 (si)
Precio de la Energía Ofertada	47,98 US$/MWh	Inicio de Montaje	31.05.2017 (si)
Puesta en Operación Comercial	30.01.2018	POC	30.01.2018 (si)

INFORMACIÓN RELEVANTE

- La Concesión Definitiva para desarrollar la actividad de generación eléctrica en la futura central de 144,48 MW, fue otorgada mediante R.M. N° 328-2017-MEM/DM del 02.08.2017.
- El 25.10.2017, el Comité de Operación Económica del Sistema Interconectado Nacional (COES) aprobó el Estudio de Operatividad de la central.
- El proyecto cuenta con Certificado de Inexistencia de Restos Arqueológicos (CIRA) aprobado para la Central Solar Fotovoltaica, caminos de acceso y L.T. 220 kV S.E. Rubí -S.E. Montalvo.
- El 31.10.2017, el COES autorizó la Conexión para las Pruebas de Puesta en Servicio de la central.
- El 10.11.2017, se conectó por primera vez la C.S.F. Rubí con una potencia de 3,5 MW.
- Durante pruebas, el 29.01.2018, generó 152,6 MW.
- El 29.01.2018, el COES aprobó la Puesta en Operación Comercial de la C.S. Rubí, con una potencia instalada nominal de 144,48 MW, a partir de las 00:00 horas del 30.01.2018.
- El monto de inversión aproximado fue de 165 MM US$, según lo indicado por la Concesionaria.

DIAGRAMA UNIFILAR





Ubicación

Vista de los paneles solares

Caseta de inversores y transformadores

Subestación Rubí

Nota. **Adaptado por OSINERGMIN (2021).**