



---

---

**UNIVERSIDAD AUTÓNOMA DE YUCATÁN**

**FACULTAD DE INGENIERÍA**

**UNIDAD DE POSGRADO E INVESTIGACIÓN**

**PREDICCIÓN DE SERIES DE TIEMPO DE POTENCIA A 24  
HORAS EN CENTRALES FOTOVOLTAICAS MEDIANTE  
REDES NEURONALES**

**TESIS**

**PRESENTADA POR:**

**ING. ELBERT DE JESÚS ALEJOS MOO**

**EN OPCIÓN AL GRADO DE:**

**MAESTRO EN INGENIERÍA**

**OPCIÓN: ENERGÍAS RENOVABLES**

**MÉRIDA, YUCATÁN, MÉXICO**

**2020**

Aunque este trabajo hubiere servido para el Examen Profesional y hubiere sido aprobado por el sínodo, sólo el autor es responsable de las doctrinas emitidas en él.

## AGRADECIMIENTOS

El autor de este documento agradece al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo financiero otorgado para la realización del presente trabajo de maestría, otorgado a partir de la beca escolar de posgrado CONACYT con número (CVU) 785709.

## RESUMEN

La tecnología solar fotovoltaica representa en la actualidad un valioso recurso energético y alternativa en muchos campos de la ciencia y vida cotidiana del ser humano. Es debido a su modularidad, que es posible crear arreglos fotovoltaicos de gran capacidad como lo son las Centrales Eléctricas Fotovoltaicas. Estas Centrales o Plantas generan electricidad debido al efecto fotovoltaico que ocurre en las celdas solares debido a la incidencia de la luz en los semiconductores de los cuales están constituidos las celdas.

Debido a que la luz que viaja desde el Sol hasta las Centrales Eléctricas Fotovoltaicas sufre alteraciones según la ubicación, ambiente y hora del día, la producción de electricidad de las centrales varía con el tiempo. Estas fluctuaciones representan un reto para los administradores de energía eléctrica, por lo que modelos de predicción o pronóstico se aplican en estos sistemas.

Con la finalidad de aportar un modelo de predicción que sea capaz de realizar pronósticos con 24 horas de anticipación con buen desempeño, este documento aborda varios modelos de redes neuronales y estadísticos. Algunos de estos modelos llamados Redes Neuronales Profundas presentan buenos desempeños en aplicaciones similares. Es por esto, que se planteó la evaluación de estos modelos en la predicción de las mediciones en Series de Tiempo de una base de datos de una Central Eléctrica Fotovoltaica, así como la validación de resultados con secciones diferentes a la utilizada para hacer los entrenamientos.

## ABSTRACT

Solar photovoltaic technology currently represents a valuable energy resource and an alternative in many fields of science and human daily life. It is due to its modularity, that it is possible to create large capacity photovoltaic arrays such as Photovoltaic Power Plants. These plants produce electricity due to the photovoltaic effect that occurs in solar cells due to the incidence of light in the semiconductors of which the cells are made.

The Sun light travels in the space to the Photovoltaic Power Plants and it causes fluctuation in the production of plants according to the location, the environment, and the time of day. These fluctuations represent a issue for electric power managers, so prediction models or forecasts are applied in these systems.

In order to provide a model that is capable of performing 24 hours prediction with high accuracy, this document addresses various statistical and neural network models. Some of these models called Deep Neural Networks perform well in similar applications. For this reason, the evaluation of these models was proposed in the prediction of measurements in Time Series from a database of a Photovoltaic Power Plant, as well as the validation of results with sections not used in the training stage.

# ÍNDICE GENERAL

AGRADECIMIENTOS.....	II
RESUMEN.....	III
ABSTRACT.....	IV
ÍNDICE GENERAL.....	V
ÍNDICE DE FIGURAS.....	VIII
ÍNDICE DE TABLAS.....	X

<b>1. INTRODUCCIÓN .....</b>	<b>1</b>
<b>1.1 Antecedentes .....</b>	<b>3</b>
<b>1.2 Planteamiento del problema .....</b>	<b>9</b>
<b>1.3 Justificación de la investigación.....</b>	<b>10</b>
<b>1.4 Objetivos .....</b>	<b>13</b>
<i>1.4.1 General .....</i>	<i>13</i>
<i>1.4.2 Específicos .....</i>	<i>13</i>
<b>2. MARCO TEÓRICO.....</b>	<b>14</b>
<b>2.1 Radiación Solar.....</b>	<b>15</b>
<i>2.2.1 Irradiancia extraterrestre .....</i>	<i>17</i>
<i>2.2.2 Irradiancia en la superficie de la Tierra.....</i>	<i>18</i>
<b>2.2 Energías Renovables .....</b>	<b>19</b>
<b>2.3 Tecnología Fotovoltaica.....</b>	<b>20</b>
<i>2.3.1 La celda solar.....</i>	<i>20</i>

2.3.2	<i>Efecto fotovoltaico</i> .....	21
2.3.3	<i>Modelo equivalente de un diodo</i> .....	22
2.3.4	<i>Efecto de la irradiancia y temperatura</i> .....	23
<b>2.4</b>	<b>Plantas Fotovoltaicas</b> .....	<b>25</b>
<b>2.5</b>	<b>Series de Tiempo</b> .....	<b>27</b>
<b>2.6</b>	<b>Teoría de embebimiento</b> .....	<b>28</b>
2.6.1	<i>Función de Autocorrelación</i> .....	32
<b>2.7</b>	<b>Métodos estadísticos</b> .....	<b>32</b>
2.7.1	<i>ARMA</i> .....	32
<b>2.8</b>	<b>Redes neuronales clásicas</b> .....	<b>33</b>
2.8.1	<i>Funciones de activación</i> .....	35
2.8.2	<i>Algoritmo de aprendizaje</i> .....	37
2.8.3	<i>Ecuación Entrada/Salida</i> .....	38
<b>2.9</b>	<b>Redes neuronales profundas</b> .....	<b>40</b>
2.9.1	<i>Breve historia del Deep Learning</i> .....	40
2.9.1	<i>Características principales</i> .....	42
2.9.1	<i>Optimizaciòn de redes profundas</i> .....	44
<b>2.10</b>	<b>Criterios estadísticos</b> .....	<b>48</b>
2.10.1	<i>Coeficiente de correlaciòn</i> .....	48
2.10.2	<i>Error Absoluto Medio Porcentual</i> .....	49
2.10.3	<i>Error Cuadrático Medio</i> .....	49
<b>3.</b>	<b>METODOLOGÍA</b> .....	<b>50</b>
<b>3.1</b>	<b>Recolección de datos</b> .....	<b>51</b>
<b>3.2</b>	<b>Minería de datos</b> .....	<b>54</b>
<b>3.3</b>	<b>Modelos de predicciòn</b> .....	<b>59</b>
<b>3.4</b>	<b>Entrenamiento de modelos inteligentes</b> .....	<b>60</b>

3.4.1	<i>Base de datos</i> .....	61
3.4.2	<i>Modelos IA</i> .....	62
3.4.3	<i>Entrenamiento</i> .....	63
3.4.4	<i>Error estadístico</i> .....	64
<b>4.</b>	<b>RESULTADOS</b> .....	<b>65</b>
4.1	<b>Procesamiento de la base de datos</b> .....	<b>65</b>
4.2	<b>Dimensión de embebimiento</b> .....	<b>67</b>
4.3	<b>Predicción con ARMA</b> .....	<b>69</b>
4.4	<b>Predicción con Redes neuronales</b> .....	<b>71</b>
4.4.1	<i>Selección de la función de transferencia</i> .....	<b>71</b>
4.4.2	<i>Determinación de la ventana de predicción</i> .....	<b>72</b>
4.4.3	<i>Selección de la función de transferencia</i> .....	<b>74</b>
4.5	<b>Predicción con Redes profundas</b> .....	<b>76</b>
4.6	<b>Comparación de resultados</b> .....	<b>78</b>
<b>5.</b>	<b>CONCLUSIONES</b> .....	<b>80</b>
	<b>Referencias</b> .....	<b>82</b>
	<b>Bibliografía</b> .....	<b>82</b>

## ÍNDICE DE FIGURAS

Fig. 2. 1. Geometría de la órbita de la Tierra y superficie del Sol para el cálculo de la Constante Solar.....	17
Fig. 2. 2. Circuito eléctrico del modelo equivalente de un diodo ideal simplificado. ....	22
Fig. 2. 3. Curvas I-V con el efecto de la irradiación incidente (a) y la temperatura (b) en una celda solar.....	24
Fig. 2. 4. Esquema de conexión de una central eléctrica fotovoltaica.....	25
Fig. 2. 5. Diagrama de interconexión de generación renovable al Sistema Eléctrico Nacional (SEN). ....	26
Fig. 2. 6. Arquitectura de red neuronal artificial multicapa. ....	35
Fig. 2. 7. Funciones de activación para neuronas artificiales.....	36
Fig. 2. 8. Arquitectura de red neuronal para modelos profundos. ....	43
Fig. 2. 9. Resultados de regresión logística en entrenamientos con algoritmos de aprendizaje profundos (Kingma y Lei, 2015). ....	46
Fig. 2. 10. Desvanecimiento del gradiente y deserción en redes neuronales profundas...	47
Fig. 3. 1. Diagrama de la metodología. Se compone de la recolección, minería de datos y aplicación de modelos de predicción.....	51
Fig. 3. 2. Gráficas de potencia eléctrica recolectadas en las plantas DREW y CVERDESOLAR en el mes de enero del 2015 (CENACE).....	52
Fig. 3. 3. Gráficas 3D de potencia eléctrica recolectadas en las plantas CVERDESOLAR y DREW en el año 2015 (CENACE).....	53
Fig. 3. 4. Diagrama de flujo para la extracción de la potencia en horas de Sol.....	57
Fig. 3. 5. Diagrama de flujo para el promedio por horas de la potencia eléctrica fotovoltaica.....	58
Fig. 3. 6. Procedimiento para el entrenamiento de los modelos. La ventana de dimensión de valores pasados se determina con el Método de Cao.....	59
Fig. 3. 7. Proceso de entrenamiento para los modelos de inteligencia artificial.....	61

Fig. 4. 1. Base de datos procesada por horas.....	66
Fig. 4. 2. Base de datos procesada por días.....	66
Fig. 4. 3. Desviación estándar de los promedios diarios de la base de datos.....	67
Fig. 4. 4. Cálculo de autocorrelaciones para el Método de Cao.....	68
Fig. 4. 5. Determinación de la dimensión de embebimiento por Cao.....	68
Fig. 4. 6. Base de datos de entrenamiento y de prueba con ARMA.....	69
Fig. 4. 7. Diferencia entre la base de datos y la predicción con ARMA.....	70
Fig. 4. 8. Comportamiento de las predicciones con la ventana de predicción.....	73
Fig. 4. 9. Diferencia entre la base de datos y la predicción con NN.....	74
Fig. 4. 10. Comportamiento de la predicción con la configuración de las neuronas en la capa oculta.....	75
Fig. 4. 11. Error entre la base de datos y la predicción con NN.....	76
Fig. 4. 12. Comportamiento de la predicción con modelos profundos.....	77
Fig. 4. 13. Error entre la base de datos y el modelo profundo.....	78

## ÍNDICE DE TABLAS

Tabla 2. 1. Ecuaciones de funciones de activación.....	37
Tabla 4. 1. Criterios estadísticos del ajuste con ARMA.....	70
Tabla 4. 2. Criterios estadísticos del ajuste con NN.....	71
Tabla 4. 3. Criterios estadísticos para la selección de ventana.....	72
Tabla 4. 4. Criterios estadísticos para la configuración del número de neuronas en la capa oculta.....	75
Tabla 4. 5. Criterios Estadísticos de la configuración de profundidad. ....	77
Tabla 4. 6. Comparación de los criterios estadísticos de los modelos ajustados.....	79

---

# Capítulo 1

## Introducción

---

En este primer capítulo se abordará las guías y directrices actuales que conllevaron al desarrollo de la tecnología solar y los modelos de pronóstico en el tiempo. Para iniciar, se puede mencionar que la humanidad extrae recursos del medio ambiente para satisfacer sus necesidades energéticas. Algunos recursos son denominados combustibles fósiles, que entre los más utilizados podemos mencionar el petróleo, carbón y gas natural, clasificados como recursos no renovables, lo que significa que eventualmente las reservas se agotarán. La producción de energía mundial utilizando estas fuentes es del 82 % (WEC, 2013). Se conoce desde hace años atrás que la combustión de recursos fósiles resulta en emisiones de gases de efecto invernadero, generando contaminación del aire y perjuicios a la biosfera (Judkins, Fulkerson, & Sanghvi, 1993).

Dado el dilema de la combustión de fósiles, se fueron incrementando las investigaciones de fuentes alternativas amigables con el ambiente como lo son: Energía solar, eólica, mareomotriz, por mencionar algunos. La energía del Sol es la fuente más abundante y disponible para uso directo (radiación solar) o indirecto (viento, biomasa, etc.) (WEC, 2013). Si consideramos el desierto del Sahara para ilustrarnos la abundancia del sol, en

esta ubicación la irradiación anual es cerca de 2350 kWh/m<sup>2</sup>. La irradiación anual total recibida por la superficie del Sahara (alrededor de 8,7 millones de km<sup>2</sup>) es casi 200 veces mayor que la demanda mundial anual de energía primaria; de hecho, la demanda mundial de energía primaria podría satisfacerse recolectando la energía solar recibida por 48,500 km<sup>2</sup> del Sahara, un área un poco más grande que Suiza, o una novena parte del área de California. Estas cifras muestran claramente que es posible satisfacer toda la demanda mundial de energía únicamente con energía solar (Quaschnig, 2005).

Una forma de aprovechamiento de la energía solar directa es mediante celdas fotovoltaicas. Las celdas se interconectan en serie y paralelo para formar módulos fotovoltaicos, y estas a su vez forman arreglos fotovoltaicos más complejos (Dunlop, 2010). Una central eléctrica fotovoltaica se compone de arreglos fotovoltaicos que suministran electricidad a alguna interconexión a red. Sin embargo, la radiación solar global y el consumo de electricidad por parte de los usuarios tienen comportamientos no lineales cuando son graficados contra el tiempo, por lo que se utilizan herramientas de pronóstico para la predicción de potencia, demanda, irradiancia solar, etc. Sin predicciones de consumo de electricidad con buena confiabilidad, los operadores de interconexión son incapaces de balancear con anticipación la demanda y la energía por suplir a los consumidores. En algunos mercados energéticos, existen penalizaciones cuando se hace incumplimientos con la energía producida y la anticipada (Leva, Dolara, Grimaccia, Mussetta, & Ogliari, 2017). Para predecir la potencia que suministra una central eléctrica fotovoltaica se aplican diversos métodos y modelos de estimación. Existen métodos estadísticos que implican calibraciones y reajustes constantes. Los modelos de inteligencia artificial presentan alternativas para la estimación de variables y pronóstico en series de tiempo.

Existe un modelo de la inteligencia artificial de reciente emergencia en el campo del Aprendizaje Automático (del inglés Machine Learning), es el aprendizaje profundo (del inglés Deep Learning). Los modelos de aprendizaje profundo se entienden como funciones de altas dimensiones que aprenden con una secuencia de transformaciones no lineales (Polson & Sokolov, 2017). Debido a su reciente popularidad, los modelos profundos de predicción en series de tiempo no han abarcado los problemas referidos en energías renovables, y dado su desempeño superior, las implicaciones de sus resultados van a impactar la literatura de manera positiva.

## 1.1 Antecedentes

A continuación, se presentarán los trabajos hallados en la literatura relacionados con métodos estadísticos, físicos o de inteligencia artificial para predecir valores futuros en series de tiempo en problemas de potencia eléctrica de plantas fotovoltaicas.

En Chu *et al.* 2015 se aplica un método inteligente y en tiempo real para predecir la energía generada por una planta fotovoltaica (FV) de 48 MWe. Este método de reorientación se desarrolla en base a esquemas de optimización de redes neuronales artificiales (NN, del inglés Neural Network) y se emplea para mejorar el rendimiento de tres modelos de predicción: un modelo físico determinista basado en técnicas de rastreo en la nube, un modelo de promedio móvil autorregresivo (ARMA), y un modelo k-ésimo vecino más cercano (kNN). Utilizando los datos de potencia medidos de la planta fotovoltaica, el rendimiento de todos los pronósticos se evalúa en términos de criterios estadísticos (sesgo promedio, error absoluto medio y error cuadrático medio). Con el

método de reenfoque, las habilidades de pronóstico de los tres modelos de referencia se incrementan significativamente para horizontes temporales de 5, 10 y 15 minutos. El estudio de Chu demuestra la efectividad del método optimizado de reencastillado para reducir los errores de aprendizaje producidos por un conjunto diverso de metodologías de pronóstico.

Algunos autores consideraron las pérdidas de potencia por sombreado parcial para el pronóstico de potencia en centrales eléctricas. Picault *et al.* 2010 presentó un método novedoso para pronosticar la producción existente de paneles fotovoltaicos en diversas condiciones ambientales. En ese enfoque, los datos de medición de campo se usan para identificar los parámetros del módulo de una vez por todas. El método propuesto simula matrices fotovoltaicas con esquemas de interconexión de módulos adaptables para reducir las pérdidas de desajuste. El modelo fue validado por los resultados experimentales tomados en una planta de 2.2 kWp, con tres esquemas de interconexión diferentes, que muestran una precisión de previsión de producción de energía confiable en condiciones de operación tanto parcialmente sombreadas como normales. Las mediciones de campo promovieron interés en el uso de configuraciones de planta alternativas en sistemas fotovoltaicos para disminuir las pérdidas de potencia en matrices de módulos.

En el trabajo de Hugo y Coimbra 2012 evaluaron varias técnicas de pronóstico que no utilizan entradas exógenas para predecir la producción de energía solar de una planta de energía fotovoltaica de seguimiento de eje único de 1 MWp que opera en Merced, California. Los datos de producción utilizados en este trabajo corresponden a la potencia promediada por hora recopilada de noviembre de 2009 a agosto de 2011. Los datos anteriores a enero de 2011 se utilizaron para entrenar a los diversos modelos de pronóstico para la potencia promediada horaria de 1 y 2 h. Los métodos estudiados en ese trabajo

son: modelo persistente, promedio móvil integrado autorregresivo (ARIMA), k-vecinos más cercanos (kNN), redes neuronales artificiales (NN) y RNA optimizados por algoritmos genéticos (GA / NN). La precisión de los modelos se determinó calculando las estadísticas de error tales como error absoluto medio (MAE), error de sesgo medio (MBE) y el coeficiente de correlación (R2) para las diferencias entre los valores pronosticados y los valores medidos para el período de enero a agosto de 2011. Ese trabajo también aborda la precisión de los diferentes métodos en función de la variabilidad de la producción de potencia, que depende en gran medida de las condiciones estacionales. Los resultados muestran que los modelos de pronóstico basados en NN funcionan mejor que las otras técnicas de predicción, y que se pueden lograr mejoras sustanciales con una optimización GA de los parámetros NN, y determinan que la precisión de todos los modelos depende fuertemente de las características estacionales de la variabilidad solar.

Existen diversos trabajos relacionados a la predicción a veinticuatro horas con series de tiempo, algunos implican modelos de inteligencia artificial y series de tiempo. Cococcioni *et al.* 2011 presenta un documento con un enfoque flexible para el pronóstico de la producción de energía en instalaciones solares fotovoltaicas, utilizando análisis de series de tiempo y redes neuronales con un día de previsión. La principal novedad del enfoque presentado es la propuesta de una herramienta para que el técnico de una instalación fotovoltaica configure correctamente el modelo de predicción de acuerdo con las características particulares. La configuración correcta tiene en cuenta la cantidad de neuronas ocultas, el número de elementos de retardo y el ancho de la ventana de entrenamiento, es decir, el número apropiado de días, antes del día previsto, empleado para el entrenamiento. La irradiación y la hora de muestreo se utilizan como variables de entrada para predecir la energía acumulada diariamente con un porcentaje de error inferior al 5 %.

Fernandez-Jimenez *et al.* 2012 presentaron un novedoso sistema estadístico de pronóstico a corto plazo para una planta fotovoltaica conectada a la red. El sistema propuesto se compone de tres módulos compuestos por dos modelos numéricos de predicción del clima y un modelo basado en redes neuronales artificiales. Los dos primeros módulos se utilizaron para pronosticar variables meteorológicas utilizadas por el tercer módulo, que fue seleccionado de un conjunto de diferentes modelos. El valor final previsto es la producción de energía por hora en la planta fotovoltaica. El horizonte de pronóstico oscila entre 1 y 39 horas, y abarca todo el día siguiente. Los autores justificaron que los valores pronosticados pueden utilizarse para determinar las horas más favorables para llevar a cabo tareas de mantenimiento en la planta y para preparar ofertas para el mercado eléctrico.

Bouzerdoum *et al.* 2013 presenta en su documento un novedoso modelo híbrido para el pronóstico de potencia a corto plazo de una planta fotovoltaica conectada a la red. El nuevo modelo combina dos métodos bien conocidos: el método de la media móvil integrada automática regresiva estacional (SARIMA) y el método de máquinas de vectores de soporte (SVM). Se usa una base de datos experimental de la potencia producida por una planta fotovoltaica de 20 kWp a pequeña escala para desarrollar y verificar la efectividad del modelo propuesto en el pronóstico a corto plazo. Las predicciones horarias de la potencia producida por la planta lo llevaron a cabo durante unos días mostrando una precisión bastante buena. También presentaron un estudio comparativo que muestra que el modelo híbrido desarrollado funciona mejor que el modelo SARIMA y SVM por separado.

Honglu Zhu *et al.* 2015 elaboraron un documento con la predicción de potencia para las plantas de energía fotovoltaicas conectadas a la red. El artículo presentó un método que

combina las ventajas de la descomposición wavelet (WD) y la red neuronal artificial (NN) para resolver el problema de predicción de potencia. Las variables meteorológicas se eligieron como entradas del modelo híbrido basado en WD y NN. La potencia de salida de la planta fotovoltaica se descompone usando WD para separar la información útil de las perturbaciones. Las NN se utilizan para construir los modelos de la potencia de salida fotovoltaica descompuesta. Finalmente, las salidas de los modelos NN se reconstruyen en la potencia pronosticada de la planta fotovoltaica. El método presentado se compara con el método de pronóstico tradicional basado en NN. Los resultados demuestran que el método descrito en el documento de Zhu *et al.* 2015 necesita menos tiempo de cálculo y tiene una mejor precisión de predicción.

En el documento de Leva *et al.* 2017 propone y analiza una red neuronal artificial para la predicción de energía de plantas fotovoltaicas en términos de su sensibilidad con respecto al tamaño de ventana de los datos de entrada. Además, la precisión del método se estudió en función de los conjuntos de datos de entrenamiento y las definiciones de error. El análisis se basó en actividades experimentales llevadas a cabo en una planta de energía fotovoltaica real acompañada de un modelo de cielo despejado. En particular, el documento trató de la predicción de energía por hora para todas las horas del día del día siguiente, en base a las 48 horas de previsión meteorológica. Esto es muy importante debido a las características predictivas solicitadas por la aplicación de red inteligente: planificación de fuentes de energía renovables, en particular dimensionamiento de sistemas de almacenamiento y mercado de energía.

En la literatura comenzaron a surgir trabajos que presentan arquitecturas de redes neuronales profundas para la predicción en series de tiempos. Xueheng Qiu et al. 2014 proponen por primera vez un conjunto de redes de creencias de aprendizaje profundo

DBN (Deep Belief Networks, del ingles) para la regresión y el pronóstico de series de tiempo. Otra contribución novedosa que plantearon fue agregar a las salidas de varios DBN un modelo de regresión de vector de soporte (SVM). Mostraron la ventaja del método propuesto en tres conjuntos de datos de demanda de carga de electricidad, un conjunto de datos de series temporales artificiales y tres conjuntos de datos de regresión sobre otros métodos de referencia.

Al paso del tiempo, surgieron modelos poco convencionales para predecir series de tiempo con arquitecturas profundas como las propuestas por Huaizhi Wang *et al.* 2017. Esos autores presentaron un novedoso método híbrido para la predicción determinística de la potencia fotovoltaica basada en la transformación wavelet (WT) y la red neuronal convolucional profunda (DCNN). WT se usa para descomponer la señal original en varias series de frecuencia. Cada frecuencia tiene mejores esquemas y comportamientos. DCNN se emplea para extraer las características no lineales y las estructuras invariantes exhibidas en cada frecuencia. Luego, un modelo probabilístico de predicción de potencia fotovoltaica que combina el método determinista propuesto y la regresión de cuantiles de columna vertebral (QR) se desarrolló originalmente para evaluar estadísticamente la información probabilística en datos de potencia fotovoltaica. Los métodos de pronóstico determinísticos y probabilísticos propuestos se aplicaron a series de datos fotovoltaicos reales recopilados de granjas fotovoltaicas en Bélgica. Los resultados numéricos presentados en los estudios de caso demuestran que los métodos propuestos exhiben la capacidad de mejorar las exactitudes de pronóstico en términos de estaciones y varios horizontes de predicción, en comparación con los modelos de pronóstico convencionales.

Ud Din y Marnerides 2017 presentaron en su documento la aplicabilidad y comparación del rendimiento de los modelos de redes neuronales profundas con retropropagación (FF-DNN) y redes neuronales recurrentes profundas (R-DNN) en el contexto corto plazo para el pronóstico de carga de electricidad. El método propuesto en ese documento se evaluó sobre base de datos reales recopilados en un período de 4 años y proporcionó predicciones con un día y una semana de anticipación. La contribución detrás de ese trabajo radica en la utilización de un procedimiento de selección de características de tiempo-frecuencia (TF) de la base de datos real que ayuda al procedimiento de regresión iniciado por los DNN antes mencionados. Los autores demostraron que el esquema introducido puede aprender adecuadamente los patrones ocultos y determinar con precisión el pronóstico de consumo de carga a corto plazo utilizando un rango de fuentes de entrada heterogéneas que no se relacionan necesariamente con la medición de la carga en sí misma sino también con otros parámetros tales como los efectos de clima, tiempo, vacaciones, carga de electricidad rezagada y su distribución durante el período. Los resultados generados revelan que el uso sinérgico del análisis de características TF con DNN permite obtener una mayor precisión capturando factores dominantes que afectan los patrones de consumo de electricidad asegurando una contribución significativa en los sistemas de energía de próxima generación y en Redes Inteligentes.

## **1.2 Planteamiento del problema**

Debida a la variabilidad de las condiciones del ambiente, el clima y la posición del Sol, la producción de electricidad en centrales eléctricas presenta irregularidades en el tiempo. Existen diversos métodos y modelos matemáticos de predicción en la literatura que

presentan resultados con gran precisión, entre ellos los métodos físicos, estadísticos y nuevas propuestas de inteligencia artificial. Cuando se trata de predecir valores de un proceso cualquiera, las ubicaciones geográficas donde se recolectan los datos influyen como una variante en el proceso, por lo que es importante determinar y validar si los modelos de predicción pueden ser utilizados en diferentes regiones. La utilización de los datos en series de tiempo recolectados en una ubicación en particular permitirá el ajuste de modelos de predicción, y la selección del método que tenga el mejor desempeño. Las redes neuronales profundas presentan mejores resultados que las redes convencionales o clásicas, lo que genera un campo de oportunidad en las series de tiempo. Dado a muy pocos antecedentes relacionados a arquitecturas profundas en series de tiempo, ésta investigación será buen punto de partida para futuros trabajos.

### **1.3 Justificación de la investigación**

La conversión de energía solar en electricidad tiene ventajas tales como: Instalación en corto tiempo, larga vida de operación, circuitos de interconexión simples, partes no móviles y generación de electricidad segura, no contaminante y de una fuente renovable (Mekki, Salhi, & Khaled, 2007).

Es debido a lo anterior, que numerosos países incrementaron el uso de energías renovables, como la solar, para construir centrales eléctricas. Ejemplo de esto, el gobierno de Japón propone el incremento de la capacidad de generación de potencia eléctrica fotovoltaica instalada a 28 GW y 53 GW para el 2020 y 2030, respectivamente. Otros países implementaron con motivo de planeación, control y monitorización de las interconexiones

a redes o microredes, así como para la estabilización y predicción de la potencia de las plantas generadoras fotovoltaicas, tales como España, que estableció el Centro de Control de Energías Renovables, para la supervisión de generadores eléctricos renovables arriba de 10 MW (Attaviriyanupap, Tokuhara, Marmioli, Tsukamoto, & Kojima, 2011).

En el caso de México, la Ley para el Aprovechamiento de las Energías Renovables y el Financiamiento de la Transición Energética (LAERFTE, 2008) establece que para el año 2024 la generación de electricidad mediante fuentes renovables deberá ser del 35%. Esta medida de política energética está establecida en la Estrategia Nacional de Energía 2014-2028, el cual enfoca los esfuerzos para impulsar el aprovechamiento de fuentes como la eólica, solar, biomasa, nuclear e hidráulica mediante el fortalecimiento de la cadena educación-ciencia-tecnología-innovación correspondientes a las actividades esenciales para construir una economía competitiva en el sector energético (Secretaría de Energía, 2012).

Para llevar a cabo la interconexión de las plantas generadoras fotovoltaicas ubicadas en México, el Centro Nacional de Control de Energía (CENACE) es el organismo que define las características de la infraestructura requerida para realizar la interconexión, como fue publicado en el DOF (Diario Oficial de la Federación, 2015). Hasta octubre del 2016, el CENACE atiende 746 solicitudes de interconexión, de los cuales se reportan 387 de tecnología fotovoltaica y 20 de estas instalaciones provienen de Yucatán, por lo cual se necesitan herramientas de predicción y control aplicables a la Península.

La generación de electricidad en plantas fotovoltaicas es fluctuante en el tiempo. La salida de potencia eléctrica en los arreglos fotovoltaicos depende de las condiciones meteorológicas, radiación solar y de carga. Como se ha demostrado en experimentos mencionados en literatura (Henan, 2013), es la variación de irradiancia solar el parámetro más relevante en la generación de electricidad en los arreglos fotovoltaicos. Debido a que

la irradiancia que incide sobre una celda fotovoltaica tiene un comportamiento no lineal frente al tiempo, se puede observar curvas en forma de campana en las gráficas intensidad-voltaje (I-V) con las mediciones de un día. Otras condiciones atmosféricas como el viento, temperatura ambiental, nubosidad y humedad también afectan a la eficiencia de los arreglos fotovoltaicos según Henan 2013.

Es debido a las variaciones en la generación eléctrica de arreglos fotovoltaicos que los administradores de red de energía eléctrica requieran constantemente ajustar modelos de predicción. En la literatura podemos encontrar los llamados métodos estadísticos de prospección, que pueden ser utilizados en la predicción de la generación de energía. Entre los modelos de análisis de regresión y series de tiempo, encontramos tales como: El modelo de regresión lineal, Regresión Lineal Múltiple, Regresión no Lineal, Modelo Autorregresivo (AR), Promedio Móvil (MA) y Series de Tiempo No-Estacionales. Estos modelos estadísticos requieren un proceso de recalibración y ajuste constantemente, que en principio tarda más tiempo en completarse, a esto se le agrega la baja precisión y la necesidad de nuevos datos históricos. Para sobrellevarlo, existe una amplia gama de modelos de inteligencia artificial relacionados al pronóstico de generación de energía: los sistemas expertos, de lógica difusa, redes neuronales artificiales, etc, que utilizan aprendizajes del problema en cuestión al sistema de energía bajo carga, y así realizar previsiones (Oudjana, Hellal, & Mahammed, 2013).

La estimación de la potencia en centrales fotovoltaicas contribuye a la administración optimizada del suministro de energía a las redes de distribución, instalación de nuevas centrales y la evolución de este tipo de tecnologías. Indirectamente, los resultados benefician al planeta con la disminución de las emisiones que producen las fuentes convencionales de energía. La predicción ayuda para la planeación, implementación y

redimensionamiento de centrales fotovoltaicas. Las predicciones a futuro permiten hacer un balance del control, estabilizarlo, hacer transacciones de potencia e interrupciones de potencia en sistemas autónomos y conectados a red (Oudjana, Hellal, & Mahammed, 2013).

## 1.4 Objetivos

### 1.4.1 General

Ajustar modelos de predicción con veinticuatro horas de anticipación para potencia eléctrica de centrales eléctricas fotovoltaicas en series de tiempo con métodos de inteligencia artificial.

### 1.4.2 Específicos

1. Predecir la potencia eléctrica de centrales eléctricas fotovoltaicas con modelos de aprendizaje profundo, redes neuronales clásicas y modelos estadísticos.
2. Obtener criterios estadísticos para las predicciones en los procesos de entrenamiento y prueba de los modelos seleccionados.
3. Elaborar un análisis comparativo de los modelos para determinar el desempeño en la serie temporal de potencia eléctrica de la planta fotovoltaica.

---

## Capítulo 2

### Marco Teórico

---

Dado que el enfoque principal de esta tesis es la predicción de potencia en centrales eléctricas fotovoltaicas, resulta fundamental definir los fenómenos que se atribuyen a la generación de potencia de los módulos fotovoltaicos, las características de los modelos de predicción actuales en la literatura y los procedimientos matemáticos implicados. La generación eléctrica fotovoltaica comprende conceptos nucleares, astronómicos, atmosféricos y de los principios de funcionamiento de los dispositivos de transformación. El camino que atraviesa un haz de luz desde su emisión hasta su captación es afectado por aspectos geométricos, físicos y químicos de los diversos medios donde se transporta. Debido que esa luz es captada por los dispositivos fotovoltaicos, se deduce que un registro en series de tiempo comprende implícitamente los fenómenos anteriormente mencionados. Incluyendo las características propias de los dispositivos fotovoltaicos frente a la radiación, la localización y su dimensión, se determina que una predicción es prácticamente compleja. Debido a que el Sol es la fuente de luz que utilizan los arreglos fotovoltaicos, se explicará a continuación los principios que rigen todo el camino de un haz de luz desde su generación hasta su captación y transformación.

## 2.1 Radiación Solar

El Sol es el punto central de nuestro sistema solar; probablemente ha estado en existencia durante 5 mil millones de años y se espera que sobreviva durante otros 5 mil millones de años. El Sol consiste en aproximadamente un 80 por ciento de hidrógeno, un 20 por ciento de helio y solo un 0.1 por ciento de otros elementos (Duffle & Beckman, 2005).

Los procesos de fusión nuclear crean el poder radiante del Sol. Durante estos procesos, cuatro núcleos de hidrógeno (protones  ${}^1\text{p}$ ) se fusionan para formar un núcleo de helio (partícula  ${}^4\alpha$ ). La partícula alfa consta de dos neutrones  ${}^1\text{n}$  y dos protones con carga positiva  ${}^1\text{p}$ . Además, esta reacción produce dos positrones  $e^+$  y dos neutrinos  $\nu_e$  y genera energía. Matemáticamente, la Ec. 2.1 describe el balance energético de fusión:



La comparación de las masas de las partículas atómicas antes y después de la reacción muestra que la masa total se reduce por la reacción. Este cambio de masa esta dado por:

$$\Delta m = 4 \cdot m({}^1\text{p}) - m({}^4\alpha) - 2 \cdot m(e^+)$$

Tomando las masas de las partículas en Quaschnig 2005, el balance de masas de la reacción resulta como:

$$\Delta m = 4 \cdot 1.00727647 \text{ u} - 4.0015060883 \text{ u} - 2 \cdot 0.00054858 \text{ u} = 0.02650263 \text{ u}$$

Por lo tanto, se observa que la masa total de todas las partículas después de la fusión es menor que la anterior. La diferencia de masa se convierte en energía  $\Delta E$ , con la relación:

$$\Delta E = \Delta m c^2 \quad \text{Ec. 2. 2}$$

y sustituyendo la constante de la velocidad de la luz  $c = 2.99792458 \cdot 10^8 \text{ m / s}$ , esta ecuación determina la energía liberada por la fusión como  $\Delta E = 3.955 \cdot 10^{-12} \text{ J} = 24.687 \text{ MeV}$ . Este proceso de fusión se repite constantemente, permitiendo al Sol generar la energía que se irradia al espacio.

Dado que el Sol pierde 4.3 millones de toneladas métricas de masa por segundo ( $\Delta m 4.3 \cdot 10^9 \text{ kg / s}$ ), es posible obtener la potencia radiante solar  $\varphi_{e,s}$  con la Ec. 2. 2:

$$\varphi_{e,s} = \Delta m c^2 = 3.845 \cdot 10^{26} \text{ W}$$

Este valor dividido por la superficie del Sol  $A_S$ , proporciona la emisión específica del Sol  $M_{e,s}$ :

$$M_{e,s} = \frac{\varphi_{e,s}}{A_S} = 63.11 \frac{\text{MW}}{\text{m}^2}$$

Esta emisión específica disminuye cuando se incrementa la distancia al Sol. El valor cerca de la superficie de la Tierra se calcula a continuación.

### 2.2.1 Irradiancia extraterrestre

La emisión específica del Sol  $M_{e,S}$  considerada en su superficie, se traslada a la órbita de la Tierra tomando la potencia radiante solar  $\varphi_{e,S}$  dividida por el área con un radio del centro del Sol a la órbita terrestre  $r_{SE}$ . Debido que la órbita de la Tierra alrededor del Sol es elíptica, se considera para estos cálculos una órbita circular promedio para determinar la irradiancia experimentada en la capa exterior de la atmósfera terrestre, como se observa en la *Fig. 2. 1*.

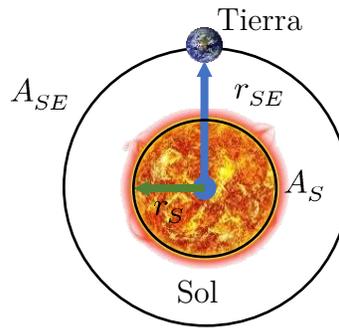


Fig. 2. 1. Geometría de la órbita de la Tierra y superficie del Sol para el cálculo de la Constante Solar.

La superficie  $A_{SE}$  de un círculo con el Sol como centro y con un radio igual a la distancia promedio desde la Tierra al centro del Sol ( $r_{SE} = 1.5 \cdot 10^8 \text{ km}$ ) recibe la misma potencia radiante total que la superficie del Sol  $A_S$ . Sin embargo, la emisión específica, o la densidad de energía medida en un metro cuadrado, es mucho más alta en la superficie del Sol que en la esfera que rodea al Sol. Con  $M_{e,S} A_S = E_e A_{SE}$  la irradiancia en la Tierra se convierte en:

$$E_e = M_{e,S} \frac{A_S}{A_{SE}} = M_{e,S} \frac{r_S^2}{r_{SE}^2}$$

Dado que la distancia entre la Tierra y el Sol no es constante, la irradiancia extraterrestre varía entre  $1325 \text{ W/m}^2$  y  $1420 \text{ W/m}^2$ . El valor promedio, llamado constante solar  $E_0$  es (Quaschnig, 2005):

$$E_0 = 1,367 \pm 2 \text{ W/m}^2$$

### *2.2.2 Irradiancia en la superficie de la Tierra*

Los valores medidos en la superficie de la Tierra son usualmente menores que la constante solar. La influencia de la atmósfera reduce la irradiancia. Algunos fenómenos que provocan la reducción de la irradiancia son (Quaschnig, 2005):

1. Reducción debido a la reflexión de la atmósfera.
2. Reducción debido a la absorción en la atmósfera (principalmente por  $O_3$ ,  $H_2O$ ,  $O_2$ , y  $CO_2$ ).
3. Reducción debido a la dispersión de Rayleigh.
4. Reducción por la dispersión de Mie.

Otros factores que influyen en la irradiancia sobre la superficie son los valores AM del Sol, así como factores climáticos como las lluvias o nubes. La irradiancia sobre la superficie es de gran importancia debido a que los dispositivos fotovoltaicos dependen de gran

medida a la irradiancia recibida, la temperatura del ambiente y factores internos de los sistemas de conversión.

## 2.2 Energías Renovables

Las fuentes de energía renovables son recursos que son inagotables dentro del horizonte temporal de la humanidad. Los tipos de energía renovables se pueden subdividir en tres áreas: energía solar, energía planetaria y energía geotérmica. La energía anual disponible respectivamente a nivel mundial es:

1. Energía solar: 3,900,000,000 PJ/año
2. Energía gravitatoria: 94,000 PJ/año
3. Energía geotérmica: 996,000 PJ/año

La energía almacenada en el viento o la lluvia, que también puede explotarse técnicamente, se origina en las conversiones de energía natural. El suministro anual de energía renovable supera en varios órdenes de magnitud la demanda mundial de energía que se muestra en la sección anterior sobre la evolución de la demanda mundial de energía. Las energías renovables pueden, teóricamente, cubrir la demanda global de energía sin ningún problema. Sin embargo, eso no significa que la transición de nuestro suministro de energía actual a los suministros de energía renovable será posible sin ningún problema. Por el contrario, los suministros de energía renovable necesitan una infraestructura totalmente diferente en comparación con lo que se ha creado durante las últimas décadas (Quaschnig, 2005).

Con mucho, el recurso de energía más grande es el Sol. Anualmente,  $3.9 \cdot 10^{24} J = 1.08 \cdot 10^{18} kWh$  de energía solar alcanza la superficie de la Tierra. Esto es aproximadamente diez mil veces más que la demanda de energía primaria mundial anual y mucho más que todas las reservas de energía disponibles en la Tierra. En otras palabras, usar una diez milésima parte de la luz solar entrante cubriría toda la demanda de energía de la humanidad (Quaschnig, 2005). De las tecnologías que usan la energía del Sol de manera directa, este trabajo está enfocado a la transformación de la radiación solar en electricidad.

## **2.3 Tecnología Fotovoltaica**

La tecnología fotovoltaica convierte la energía solar a electricidad con el uso de ciertas propiedades de algunos materiales semiconductores. Los dispositivos fotovoltaicos usan obleas, hechos típicamente de silicio cristalino, que son sensibles a la luz solar y producen una corriente directa cuando son expuestos a la luz. Las celdas solares se combinan para formar módulos fotovoltaicos como se explica en los apartados siguientes (Dunlop, 2010).

### *2.3.1 La celda solar*

Una celda fotovoltaica es un dispositivo semiconductor que convierte la radiación solar en electricidad de corriente directa. Debido a que la fuente de radiación es el Sol, son llamadas celdas solares. Las celdas solares son los elementos básicos que construyen los bloques de los módulos fotovoltaicos.

Las celdas fotovoltaicas son hechas de materiales semiconductores que son materiales que exhiben propiedades tanto aislantes y conductores cuando son sometidos a diferentes condiciones. A temperaturas ambientales, la conductividad eléctrica de un semiconductor esta entre aislante y conductor. Algunos semiconductores incluso pueden producir un voltaje o exhibir un cambio en la conductividad cuando son expuestos a la luz (Dunlop, 2010).

### *2.3.2 Efecto fotovoltaico*

El proceso básico en el cual una celda solar convierte la luz en electricidad es conocido como el efecto fotovoltaico. Este efecto se refiere al movimiento de electrones a través de un material cuando este absorbe fotones con ciertos niveles de energía. Los fotones contienen diferentes niveles de energía dependiendo de sus longitudes de onda, los fotones con energías elevadas se asocian con longitudes de ondas cortas (alta frecuencia). Los fotones de luz transfieren su energía a los electrones que se encuentran en la superficie del material. Los electrones con energía suficiente escapan de sus átomos y contribuyen una corriente eléctrica. Debido al resultado eléctrico, el efecto fotovoltaico es conocido también como efecto fotoeléctrico. Una celda fotoeléctrica es delgada, una oblea plana que consiste en una unión P-N. Cuando la unión P-N es iluminada, la absorción de fotones con altas energías produce el movimiento de los electrones extras en el material. Los electrones con energía potencial ganada están disponibles para realizar un trabajo antes de retornar a un estado de baja energía (Dunlop, 2010).

### 2.3.3 Modelo equivalente de un diodo

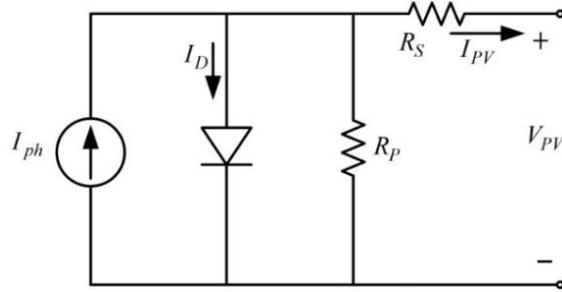


Fig. 2. 2. Circuito eléctrico del modelo equivalente de un diodo ideal simplificado.

Una celda fotovoltaica puede ser modelada con un circuito equivalente formado por una fuente de corriente con un diodo en paralelo, una resistencia en serie y una en paralelo, como se ilustra en la Fig. 2. 2. Las características eléctricas del circuito son:  $V_{PV}$  voltaje entre los bornes de la celda,  $I_{PV}$  corriente que entrega la celda,  $I_{ph}$  corriente fotogenerada producida por la energía radiante,  $I_D$  intensidad del diodo,  $R_s$  resistencia de carga del material semiconductor y  $R_p$  como la resistencia debido a impurezas y no linealidades.

La corriente  $I_{PV}$  de una celda solar está dada por:

$$I_{PV} = I_{ph} - I_D \left( e^{\frac{V_{PV} + I_{PV} R_s}{V_t}} \right) - \frac{V_{PV} + I_{PV} R_s}{R_p} \quad \text{Ec. 2. 3}$$

donde  $I_{ph}$  es la corriente de saturación inversa producida por la recombinación de portadores,  $V_t$  es el voltaje térmico dado por la expresión siguiente:

$$V_t = \frac{nkT}{e} \quad \text{Ec. 2. 4}$$

donde  $V_t$  es la expresión del voltaje térmico,  $T$  representa la temperatura de celda en grados kelvin,  $k$  es la constante de Boltzmann,  $e$  es la carga del electrón y  $n$  es el factor de idealidad del diodo (Sahnoun, Romero, Carmona, & Gomand, 2013).

De la ecuación Ec. 2. 3 se grafica la curva I-V de una celda fotovoltaica con los datos del fabricante. Dado que  $I_{PV}$  es dependiente de la corriente  $I_0$  (producida por la luz recibida) y por la Temperatura, se infiere que los fenómenos meteorológicos y de disponibilidad del recurso solar impactan a las celdas fotovoltaicas. Es decir, las condiciones ambientales de la instalación son reflejados directamente en la potencia eléctrica de la celda solar. Esto puede quedar registrado en una serie de tiempo, para investigar las fluctuaciones de los fenómenos climáticos de un sistema fotovoltaico.

#### *2.3.4 Efecto de la irradiancia y temperatura*

Como se analiza en la Ec. 2. 3, la tensión y corriente generada en una celda depende directamente de la iluminación recibida. En la Fig. 2. 3. se presentan diferentes curvas características con la finalidad de analizar los efectos de la irradiancia incidente en una celda solar y los provocados por su exposición a la temperatura de ambiente. Los valores típicos en una curva Intensidad-Voltaje (I-V) son: Voltaje de circuito abierto  $V_{OC}$  (carga cero, intensidad cero) e Intensidad de corto circuito  $I_{CC}$  (carga infinita, voltaje cero).

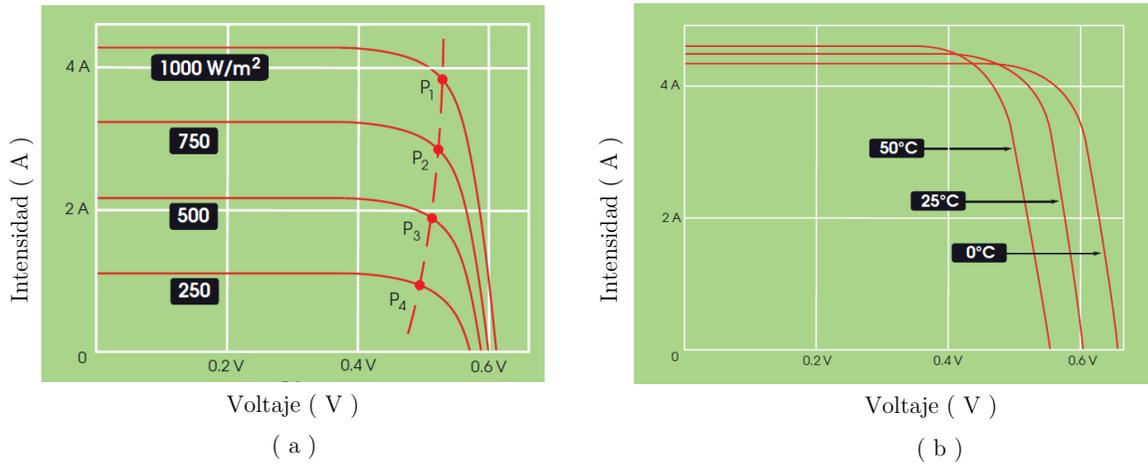


Fig. 2. 3. Curvas I-V con el efecto de la irradiación incidente (a) y la temperatura (b) en una celda solar.

Se observa que la curva superior en la Fig. 2. 3. (a), etiquetada como  $1000 \text{ W/m}^2$ , se refiere a la irradiancia estandar. Las otras tres curvas confirman que a medida que el nivel de irradiación disminuye, la corriente cae en proporción. Cada uno tiene su propio punto de potencia máxima (MPP), etiquetado  $P_1$  a  $P_4$ .

El efecto principal de la temperatura sobre la curva característica I - V de una celda es una reducción en el voltaje de circuito abierto, ilustrado por la Fig. 2. 3. (b). Además de la curva a temperatura estándar de  $25^\circ \text{C}$ , se agrega dos curvas más para  $0$  y  $50^\circ \text{C}$ . La tensión del circuito abierto cambia en aproximadamente  $0.1 \text{ V}$  entre estos extremos, lo que corresponde a  $0.33\%$  por  $^\circ \text{C}$ . Cuando el coeficiente de temperatura es negativo, el voltaje disminuye a medida que la temperatura aumenta. Hay un efecto mucho más pequeño en la corriente de cortocircuito. En general, la celda pierde potencia a temperaturas elevadas, un efecto más grave con silicio cristalino que la mayoría de los otros tipos de celdas solares (Lynn, 2010).

## 2.4 Plantas Fotovoltaicas

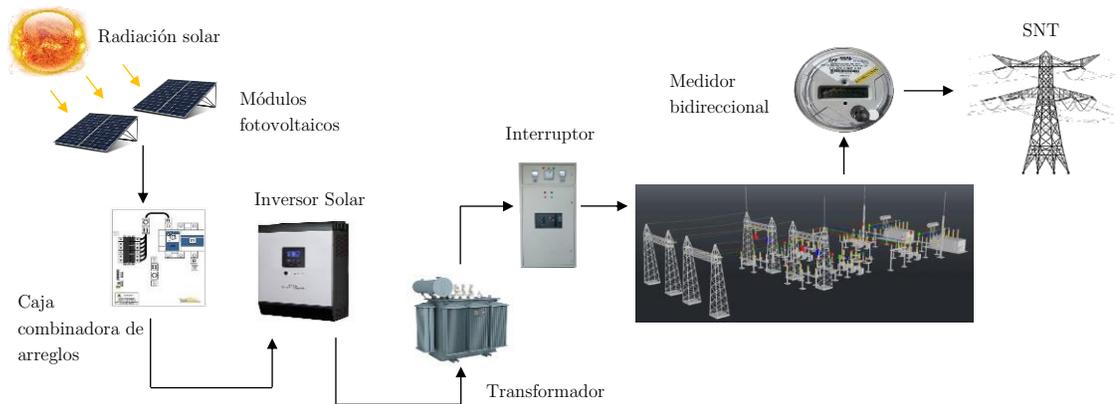


Fig. 2. 4. Esquema de conexión de una central eléctrica fotovoltaica.

Una central eléctrica fotovoltaica conectada a la red consta de módulos fotovoltaicos, inversores, unidades de acondicionamiento de energía, protecciones y equipos de medición (Shiva Kumar & Sudhakar, 2015). El funcionamiento básico de una central fotovoltaica consiste en un arreglo de módulos fotovoltaicos que convierten la luz en electricidad y se inyectan a la red pública. Debido a la generación de corriente directa, la electricidad generada por los módulos debe ser pasada a inversores, elevar el voltaje con transformadores hasta la tensión de suministro de la red y finalmente conectada con sus debidas protecciones y medidores, como en la Fig. 2. 4.

En México, la SENER, mediante la resolución RES/119/2012 indica que la comisión reguladora de energía debe expedir las reglas generales de interconexión al sistema eléctrico nacional para generadores o permisionarios con fuentes de energías renovables o cogeneración eficiente (COMISIÓN REGULADORA DE ENERGÍA, 2016).

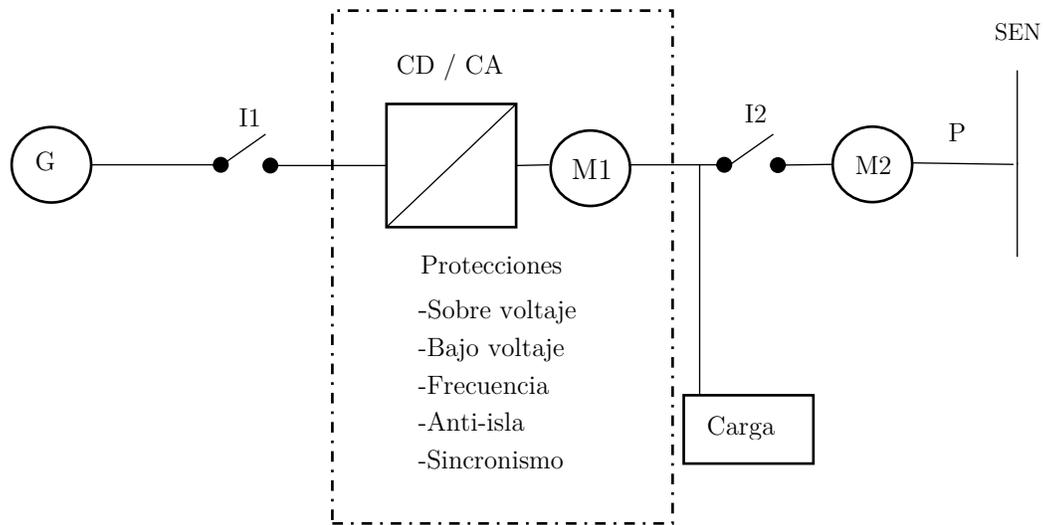


Fig. 2. 5. Diagrama de interconexión de generación renovable al Sistema Eléctrico Nacional (SEN).

El esquema básico de interconexión que presenta la resolución RES/119/2012 se muestra en la Fig. 2. 5. La G representa a la unidad de generación, que puede ser eólica o fotovoltaica. Los interruptores I1 y I2 encierran al sistema de protección, donde se localizan los dispositivos de inversión y transformación. El medidor de la generación total M1 puede venir integrado al equipo de producción, pero para fines estadísticos se requiere la instalación de este medidor a la salida de la fuente de generación conforme lo indica la RLAERFTE. El medidor M2 corresponde al medidor para facturación y tiene las siguientes características: Medidor electrónico de clase 15 de 100 amperes o clase 30 de 200 amperes, según corresponda a la carga y tipo de medición del cliente; 1, 2 o 3 fases y

rango de 120 a 480 Volts, base tipo “S”, formas 1S, 2S, 12S o 16S. La clase de exactitud es de 0.5% de acuerdo a la Especificación de CFE GWH00-78, con medición de kWh bidireccional (CFE, 2010).

Los medidores y los transformadores de instrumento destinados a la facturación deben ser instalados en el punto de interconexión. Los medidores deben contar con acceso remoto mediante un canal dedicado. Las características del acceso remoto serán definidas según cada proyecto por el área de medición del suministrador. Se debe cumplir con la especificación CFE G0000-48 “Medidores Multifunción para Sistemas eléctricos” y con las normas NRF-027-CFE y NRF-026-CFE para los transformadores de instrumento.

Entre las especificaciones de la norma CFE G0000-48 se encuentra que la medición debe ser capaz de grabar en memoria masiva los parámetros de calidad de la energía, tales como: decrementos repentinos de la tensión (Sags), incremento repentino de tensión (Swells), interrupciones, parpadeo, forma de onda con límites programables y captura de forma de onda con muestreo de al menos 128 valores por segundo.

## **2.5 Series de Tiempo**

Una serie de tiempo es un conjunto de observaciones  $x_t$ , cada una de las cuales se registra en un tiempo específico  $t$ . Una serie de tiempo discreta es aquella en la que el conjunto  $T_0$  de tiempos en el que se realizan las observaciones es un conjunto discreto, como es el caso, por ejemplo, cuando las observaciones se realizan a intervalos de tiempo. Las series de tiempo continuo se obtienen cuando las observaciones se registran continuamente

durante un intervalo de tiempo, por ejemplo, cuando  $T_0 = [0,1]$  (Brockwell & Davis, 2002).

Una parte importante del análisis de una serie temporal es la selección de un modelo de probabilidad adecuado (o clase de modelos) para los datos. Para tener en cuenta la naturaleza posiblemente imprevisible de las observaciones futuras, es natural suponer que cada observación  $x_t$  es un valor realizado de una determinada variable aleatoria  $X_t$ .

Un modelo de series de tiempo para datos observados  $\{x_t\}$  es una especificación de la unión de distribuciones (o posiblemente solo los promedios y covariancias) de una secuencia aleatoria de variables  $\{X_t\}$  de los cuales  $\{x_t\}$  son postulados.

Para este estudio, fue tomada la serie de tiempo de potencia eléctrica que registran los equipos de medición establecidos por las normas de la CFE y CENACE. Por lo tanto, las observaciones  $x_t$  son correspondientes a las mediciones del instrumento de facturación anteriormente mencionado.

## 2.6 Teoría de embebimiento

Al intentar construir un modelo de predicción para una serie temporal, es importante poder caracterizar la dimensión de incrustación de la serie. La dimensión de inserción define el número de valores previos de la serie que van a determinar el siguiente. A continuación se explicará el Método de Cao para determinar la ventana mínima de embebimiento (Cao, 1997).

La dimensión de inserción es equivalente al orden de una ecuación diferencial que definiría la serie con respecto a los valores previos. No necesariamente nos dice qué valores son

importantes, solo cuántos. En el caso simple, se supone que los valores más recientes determinan el siguiente. Una serie temporal en la cual los dos valores anteriores son suficientes para determinar el actual (usando el ejemplo de una onda sinusoidal: se necesita el último valor para saber en qué parte de la onda se encuentra posicionado, y el valor anterior para saber si el valor está subiendo o bajando de la onda) tiene una dimensión de inserción de dos.

El método utilizado en el presente trabajo para calcular la dimensión de embedding es el llamado método de Cao o de falsos vecinos cercanos promediados. Se supone que se tiene una serie de tiempo:

$$x_1, x_2, \dots, x_N$$

Los vectores de tiempo de retardo pueden ser reconstruidos como sigue:

$$y_i(d) = (x_i, x_{i+\tau}, \dots, x_{i+(d-1)\tau}),$$

$$i = 1, 2, \dots, N(d-1)\tau$$

Donde  $d$  es la dimensión de inserción y  $\tau$  es el tiempo de retardo. Se nota que  $y_i(d)$  significa el  $i$ -ésimo vector reconstruido con dimensión de inserción  $d$ . Continuando se define:

$$a(i, d) = \frac{\|y_i(d+1) - y_{n(i,d)}(d+1)\|}{\|y_i(d) - y_{n(i,d)}(d)\|},$$

$$i = 1, 2, \dots, N - d\tau,$$

donde  $\|\cdot\|$  es una medida de distancia Euclidiana y está dado por la norma máxima, i.e.,

$$\|y_k(m) - y_l(m)\| = \max_{0 \leq j \leq m-1} |x_{k+j\tau} - x_{l+j\tau}|;$$

$y_i(d+1)$  es el  $i$ -ésimo vector reconstruido con dimensión de inserción  $d+1$ , i.e.,

$$y_i(d+1) = (x_i, x_{i+\tau}, \dots, x_{i+d\tau});$$

$n(i, d)$  ( $1 \leq n(i, d) \leq N - d\tau$ ) es un entero tal que  $y_{n(i, d)}(d)$  es el vecino más cercano de  $y_i(d)$  en el espacio de fase  $d$  dimensional en el sentido de distancia  $\|\cdot\|$  que se definió con anterioridad.  $n(i, d)$  depende de  $i$  y  $d$ .

Si  $d$  es calificado como una dimensión de inserción por el teorema de embedding o de Takens (Takens, 1981), entonces 2 puntos cualquiera los cuales permanecen cercanos en el espacio reconstruido  $d$ -dimensional estarán todavía cercanos en el espacio reconstruido  $(d+1)$ -dimensional. Tal par de puntos son llamados verdaderos vecinos, de otra manera, serán llamados falsos vecinos.

Inserción perfecta significa que no existen falsos vecinos. De la definición de  $a(i, d)$ , uno puede ver que el valor límite puede ser determinado por la derivada de la señal subyacente, por lo tanto, para diferentes puntos de fase  $i$ ,  $a(i, d)$  puede tener diferentes valores límites, al menos en principio. Además, diferentes datos de series de temporales pueden tener diferentes valores límite. Esto implica que es muy difícil y casi imposible obtener un apropiado y razonable valor límite independiente de la dimensión  $d$  y los puntos de cada trayectoria, así como de los datos considerados de la serie de tiempo.

Para evitar el problema anterior, en su lugar se define la cantidad siguiente, i.e., el valor promedio de todos los  $a(i, d)$ ,

$$E(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} a(i, d)$$

$E(d)$  es dependiente solo de la dimensión  $d$  y del retardo  $\tau$ . Para investigar esta variación de  $d$  a  $d + 1$ , se define:

$$E1(d) = E(d + 1)/E(d)$$

Se observa que  $E1(d)$  no cambia cuando  $d$  es mayor que algún valor  $d_0$ . Si la serie de tiempo viene de un atractor. Entonces  $d_0 + 1$  es la mínima dimensión de embedding que se puede establecer.

El parámetro  $\tau$  es necesario y debe darse antes de que la dimensión mínima de inserción (embedding) se determine numéricamente, independientemente de los métodos que se utilicen. A pesar de que en principio la dimensión de embedding es independiente del tiempo de retardo  $\tau$ , la mínima dimensión de embedding es dependiente de  $\tau$  en práctica. Diferentes valores de  $\tau$  pueden conducir a diferentes dimensiones mínimas de embedding, especialmente para series de tiempo continuas (para series de tiempo de mapas de tiempo discreto, la mejor elección de  $\tau$  es 1).

Se puede entender fácilmente que una buena elección de  $\tau$  puede disminuir la dimensión de inserción mínima que es necesaria para la reconstrucción del espacio de fase. Esta es también una de las razones por las que generalmente se elige un retardo de tiempo tal que cada componente de los vectores reconstruidos (lineales o no lineales) es independiente en la medida de lo posible. Los métodos para elegir tal  $\tau$  son, por ejemplo, el método de función de autocorrelación y de información mutua.

### 2.6.1 Función de Autocorrelación

Una forma de elegir el tiempo de rezago se basa en el cálculo de la función de autocorrelación muestral (Small, 2005). Esta se define para una serie de tiempo escalar  $x_n$  de  $N$  muestras como:

$$\rho(\tau) = \frac{\sum_{n=1}^N (x_{n+\tau} - \bar{x})(x_n - \bar{x})}{\sum_{n=1}^N (x_n - \bar{x})^2}$$

Donde  $\bar{x}$  denota la media muestral de la serie  $x_n$ . El criterio para la selección de  $\tau_e$  es elegir el tiempo de rezago donde  $\rho(\tau)$  alcance su primer cero. De esta forma  $x_n$  y  $x_{n+\tau_e}$  estarán linealmente incorrelacionados.

## 2.7 Métodos estadísticos

### 2.7.1 ARMA

Para describir comportamientos futuros de una serie de tiempo, existen métodos estadístico como el ARMA. Este modelo es conocido como modelo autorregresivo de media móvil (en inglés AutoRegressive Moving Average models). Para una serie de tiempo  $\{X_t\}$ , se considera un proceso ARMA si la serie es estacionaria y satisface para cada  $t$ :

$$X_t - \phi X_{t-1} = Z_t - \theta Z_{t-1}$$

donde  $\{Z_t\} \sim WN(0, \sigma^2)$  y  $\phi + \theta \neq 0$ .

usando el operador de desplazamiento hacia atrás  $B$ , puede describirse como:

$$\phi(B)X_t = \theta(B)Z_t$$

donde  $\phi(B)$  y  $\theta(B)$  son filtros lineales de:

$$\phi(B) = 1 - \phi B \text{ y } \theta(B) = 1 + \theta B$$

respectivamente.

Expandiendo a las series, se reduce el proceso MA como:

$$X_t = Z_t + (\phi + \theta) \sum_{j=0}^{\infty} \phi^{j-1} Z_{t-j} \tag{Ec. 2. 5}$$

La Ec. 2. 5 determina un valor futuro  $X_t$  mediante datos pasados establecidos por  $Z_t$ ,  $\phi$  y  $\theta$ . Los procesos estacionarios y no estacionarios se explican a profundidad en Brockwell y Davis 2002 (Brockwell & Davis, 2002).

## 2.8 Redes neuronales clásicas

Una red neuronal artificial (NN, por sus siglas en inglés) es un conjunto interconectado de unidades de procesamiento que utiliza técnicas matemáticas y computacionales para

resolver problemas a partir de datos complicados, imprecisos o faltantes. Cada una de estas unidades se denomina perceptrón o neurona artificial. Una neurona tiene un peso entrante, un sesgo y una salida dada por la función de transferencia de la suma de las entradas. La función de la neurona de salida se puede expresar matemáticamente como:

$$u(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^n f(w_i x_i + b) \quad \text{Ec. 2. 6}$$

donde  $u(\mathbf{x}, \mathbf{w})$  es la salida de la neurona,  $w_i$  es el peso sináptico,  $x_i$  es la información de entrada y  $b$  es el valor de sesgo (Fausett, 1994).

Una NN generalmente se organiza en tres capas: entrada, oculta y de salida, como se muestra en la Fig. 2. 6. El entrenamiento de una NN se puede dividir en dos fases: El primer paso consiste en actualizar los valores de activación neuronal con un algoritmo de aprendizaje, la segunda fase actualiza los pesos para minimizar el error de función que mide la diferencia entre la salida deseada y la real.

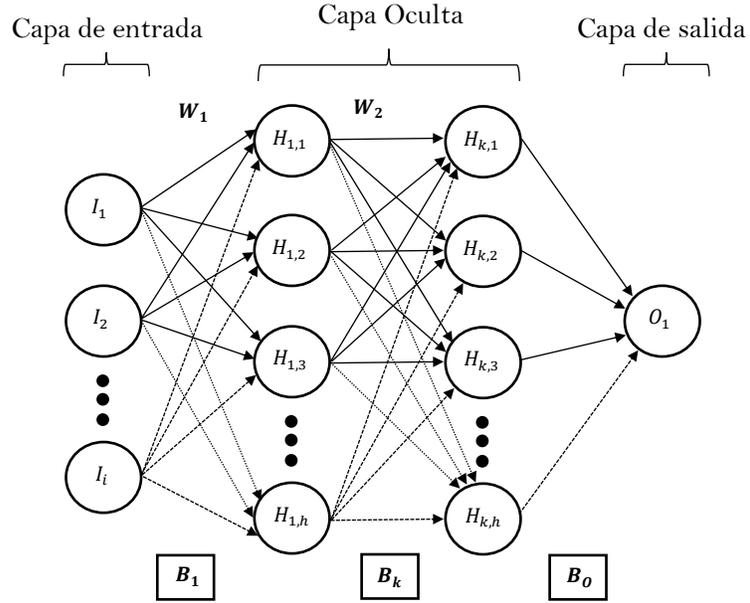


Fig. 2. 6. Arquitectura de red neuronal artificial multicapa.

### 2.8.1 Funciones de activación

La suma del producto de los pesos con las entradas de la neurona se evalúa mediante una función de activación o transferencia. Las funciones de activación populares son la Lineal, Sigmoide y Tangente hiperbólica. Estas funciones de activación son ampliamente utilizadas y son las más populares en la literatura. Existen diversas funciones de activación, tantas como el usuario desee experimentar crean nuevas funciones, cuadráticas, cúbicas entre otras.

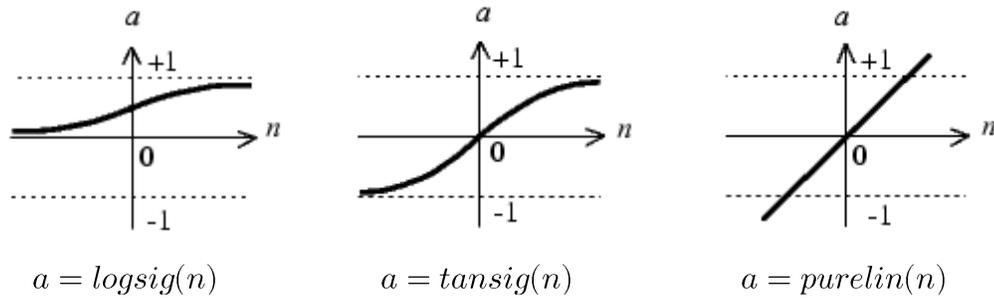


Fig. 2. 7. Funciones de activación para neuronas artificiales.

En las neuronas con función lineal, si la suma de las señales de entrada es menor que un límite inferior, la activación se define como 0 o -1. Si dicha suma es mayor o igual que el límite superior, entonces la activación es 1. Si la suma de entrada está comprendida entre ambos límites, superior e inferior, entonces la activación se define como una función lineal de la suma de señales de entrada. Si el rango de la función de transferencia no sale de algún límite, independientemente de lo grande que sea la variable de entrada, entonces la función es llamada squash. Una función squash muy popular es la función sigmoide o función logística. Es de pensarse que la función de activación sigmoide determina la ganancia no lineal de la neurona. Esta ganancia se puede definir como la razón de cambio en la señal de salida con respecto a la variación en el valor de entrada. Nótese que, en los valores pequeños de la variable de entrada, el cambio de la señal de salida es alto, mientras que, para un valor de entrada grande, el cambio en la salida es mínimo. Este tipo de funciones no lineales resuelven problemas de saturación y ruido muy común en las redes neuronales (Espadas, 2002). La *Tabla 2. 1* resume las funciones de activación para redes neuronales.

Tabla 2. 1. Ecuaciones de funciones de activación.

<i>Función de activación</i>	<i>Ecuación</i>	
<i>Lineal</i>	$f(x) = x$	Ec. 2. 7
<i>Sigmoide</i>	$\frac{1}{1 + e^{-x}}$	Ec. 2. 8
<i>Tangete hiperbólico</i>	$f(x) = \frac{2}{1 + e^{-2x}} - 1$	Ec. 2. 9

### 2.8.2 Algoritmo de aprendizaje

El gradiente descendente es uno de los algoritmos de aprendizaje más populares para realizar la optimización y, con mucho, la forma más común de optimizar las redes neuronales clásicas. Matemáticamente se expresan como:

$$\theta_{t+1} = \theta_t - \eta * \nabla_{\theta} * J(\theta)$$

dónde  $\theta$  son los pesos,  $\eta$  la tasa de aprendizaje y  $\nabla$  el gradiente del costo  $J$ . Se actualiza los parámetros en la dirección de los gradientes y con la tasa de aprendizaje se determina qué tan grande es la actualización que se realiza. El descenso del gradiente garantiza converger al mínimo global para superficies de error convexo y a un mínimo local para superficies no convexas (Tziu Tzib, y otros, 2016).

### 2.8.3 Ecuación Entrada/Salida

El ajuste del modelo NN determina una arquitectura I-H-O (número de neuronas en la capa de entrada-oculta-salida) para formular una ecuación matemática que estime una predicción de potencia ( $P_{SIM}$ ) a partir de las variables de entradas de la red. Los pesos, sesgos y funciones de activación deben ser definidas. La serie de tiempo se fracciona y se redimensiona en una matriz  $N \times M$ , donde el ancho es determinado por el Método de Cao (neuronas I). Cada valor de la ventana se inserta en su respectiva neurona en la capa de entrada, sin tener que realizar ningún cómputo.

Para el cálculo de  $u_o$ , que es la salida de una neurona en la capa oculta, se procede a multiplicar los pesos de cada entrada con el valor de cada neurona en la capa da entrada, matemáticamente:

$$u_o = W_{1(o,1)} * X_1 + W_{1(o,2)} * X_2 + \dots + W_{7(o,7)} * X_7 + b_{1o}$$

donde  $W_1$  es la del arreglo de pesos entre la capa de entrada y la oculta,  $X$  el arreglo con las variables de entrada y  $b_1$  el arreglo de sesgos de la capa oculta.

Suponiendo que  $u_o$  fuera evaluada con la función tansig, para calcular la salida de la neurona denotada como  $h_o$  se procede a evaluar el resultado de  $u_o$  tomando la *Ec. 2. 9.*

$$h_o = f_1(u_o)$$

$$h_o = \text{tansig}(u_o)$$

$$h_o = \frac{2}{1 + e^{-2u_o}} - 1$$

Estableciendo que la última capa tiene una neurona, relacionada a una predicción (paso siguiente), su entrada  $u_S$  esta dada por la sumatoria del producto de los pesos  $W_2$  (arreglo de pesos entre la capa oculta y de salida) por la salida correspondiente de las neuronas de la capa oculta:

$$u_S = h_1 * W_{2_1} + h_2 * W_{2_2} + \dots + h_O * W_{2_O} + b_2$$

donde  $O$  representan las neuronas de la capa oculta. Como la función de activación de la capa de salida es purelin por simplicidad, la predicción  $P_{SIM}$  de la arquitectura final está dada por:

$$P_{SIM} = f_2(u_S)$$

$$P_{SIM} = \text{purelin}(u_S)$$

$$P_{SIM} = h_1 * W_{2_1} + h_2 * W_{2_2} + \dots + h_O * W_{2_O} + b_2$$

Resolviendo con sumatorias, la ecuación final está dada como:

$$P_{SIM} = \sum_{O=1}^N \left[ W_{2_O} * \left( \frac{2}{1 + e^{-2(\sum_{E=1}^C (W_{1(O,E)} * X_E) + b_{1_O})}} - 1 \right) \right] + b_2 \quad \text{Ec. 2. 10}$$

## 2.9 Redes neuronales profundas

### *2.9.1 Breve historia del Deep Learning*

La historia de Deep Learning (aprendizaje profundo) se remonta a 1943, cuando Walter Pitts y Warren McCulloch crearon un modelo de computadora basado en las redes neuronales del cerebro humano. Usaron una combinación de algoritmos y matemáticas que llamaron "lógica de umbral" para imitar el proceso de pensamiento. Desde ese momento, Deep Learning ha evolucionado constantemente, con solo dos interrupciones significativas en su desarrollo, popularmente conocidos como "Inviernos de la Inteligencia Artificial".

A Henry J. Kelley se le atribuye el mérito de haber desarrollado los principios básicos de un Modelo de propagación hacia atrás en 1960. En 1962, Stuart Dreyfus desarrolló una versión más simple basada únicamente en la regla de la cadena. Si bien el concepto de retro-propagación sí existía a principios de la década de 1960, era torpe e ineficiente, y no sería útil hasta 1985.

Los primeros esfuerzos en el desarrollo de algoritmos de Aprendizaje Profundo vinieron de Alexey Grigoryevich Ivakhnenko (desarrollando el Método Grupal de Manejo de Datos) y Valentin Grigor'evich Lapa (autor de Técnicas Cibernéticas y de Pronóstico) en 1965. Utilizaron modelos con funciones de activación polinómicas (ecuaciones complicadas), que fueron analizados estadísticamente. De cada capa, las mejores características elegidas estadísticamente se reenviaron a la siguiente capa (un proceso lento y manual).

La propagación hacia atrás, el uso de errores en el entrenamiento de modelos de aprendizaje profundo, evolucionó significativamente en 1970. Esto fue cuando Seppo Linnainmaa escribió su tesis de maestría, incluyendo un código FORTRAN para la propagación hacia atrás. Desafortunadamente, el concepto no se aplicó a las redes neuronales hasta 1985. Esto fue cuando Rumelhart, Williams y Hinton demostraron que la propagación hacia atrás en una red neuronal podía proporcionar representaciones de distribución "interesantes". Filosóficamente, este descubrimiento sacó a la luz la pregunta dentro de la psicología cognitiva de si la comprensión humana se basa en la lógica simbólica (computacionalismo) o representaciones distribuidas (conexionismo).

En 1995, Dana Cortes y Vladimir Vapnik desarrollaron la máquina de vectores de soporte (un sistema para mapear y reconocer datos similares). El LSTM (memoria larga a corto plazo) para redes neuronales recurrentes fue desarrollada en 1997 por Sepp Hochreiter y Juergen Schmidhuber.

El siguiente paso evolutivo significativo para Deep Learning tuvo lugar en 1999, cuando las computadoras comenzaron a ser más rápidas en el procesamiento de datos y se desarrollaron GPU (unidades de procesamiento de gráficos). Procesamiento más rápido, con GPU procesando imágenes, velocidades computacionales aumentadas en 1000 veces durante un lapso de 10 años. Durante este tiempo, las redes neuronales comenzaron a competir con las máquinas de vectores de soporte. Mientras que una red neuronal puede ser lenta en comparación con una máquina de vectores de soporte, las redes neuronales ofrecen mejores resultados utilizando los mismos datos. Las redes neuronales también tienen la ventaja de seguir mejorando a medida que se agregan más datos de entrenamiento.

Alrededor del año 2000, apareció “The Vanishing Gradient Problem”. Se descubrió que las capas superiores no aprendían las "características" formadas en las capas inferiores, ya que ninguna señal de aprendizaje llegaba a estas capas. La fuente del problema resultó ser ciertas funciones de activación. Varias funciones de activación acotan su entrada, reduciendo el rango de salida de una manera un tanto caótica. Esto produjo grandes áreas de entrada mapeadas en un rango extremadamente pequeño. En estas áreas de entrada, un cambio grande se reducirá a un pequeño cambio en la salida, dando como resultado un desvanecimiento del gradiente.

En el pasado, muchos investigadores no pudieron entrenar redes neuronales con múltiples capas, las redes a menudo estaban atrapadas en mínimos locales cuando se inicializaban con pesos aleatorios, dando lugar al sobreentrenamiento.

Para 2011, la velocidad de las GPU había aumentado significativamente, lo que permitió entrenar redes neuronales convolucionales "sin" el entrenamiento previo capa por capa. Con la mayor velocidad de computación, se hizo obvio que Deep Learning tenía ventajas significativas en términos de eficiencia y velocidad. La evolución de los modelos profundos y algoritmos de aprendizaje fueron de la mano a la tecnológica. Schmidhuber 2015 recopila los modelos y menciona el momento en que las arquitecturas se comenzaron a distinguir entre profundas y superficiales (Schmidhuber, 2015).

### *2.9.1 Características principales*

Las redes profundas de propagación hacia adelante, también denominadas red de perceptrones multicapa (MLP), son los modelos de aprendizaje profundo por excelencia.

El objetivo de una red neuronal profunda es aproximar alguna función  $f$ . Estos modelos se denominan de propagación hacia adelante porque la información fluye a través de la función que se evalúa desde  $x$ , a través de los cálculos intermedios utilizados para definir  $f$ , y finalmente a la salida  $y$ .

No hay conexiones de retroalimentación en las cuales las salidas del modelo se realimentan a sí mismas. Cuando las redes neuronales se extienden para incluir conexiones de retroalimentación, se llaman redes neuronales recurrentes.

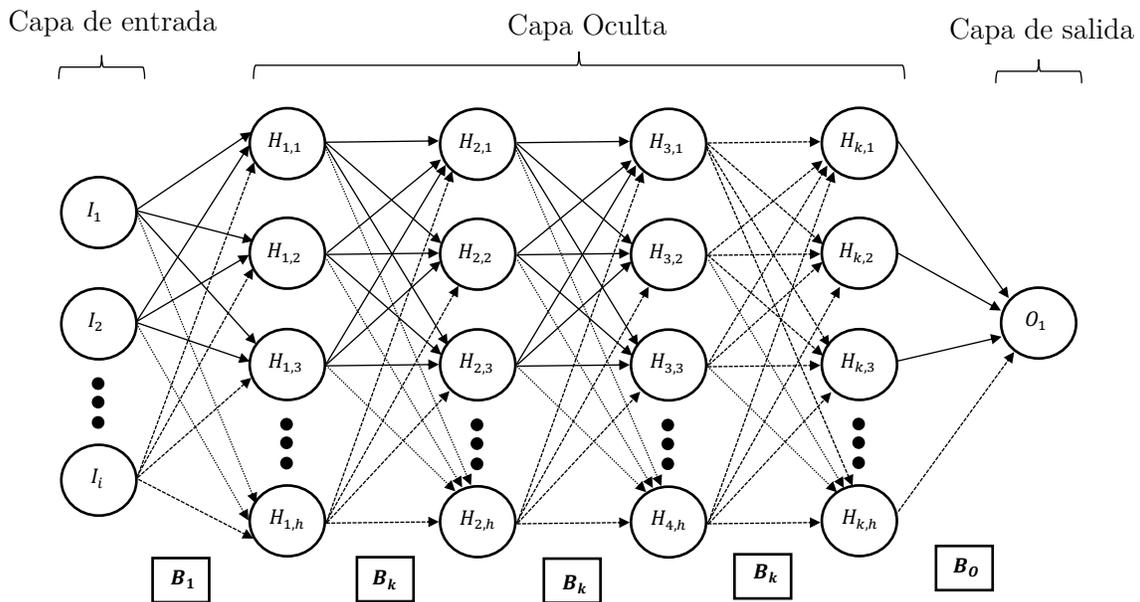


Fig. 2. 8. Arquitectura de red neuronal para modelos profundos.

Las redes neuronales se llaman redes porque típicamente se representan componiendo muchas funciones diferentes. El modelo está asociado con un gráfico acíclico dirigido que describe cómo se componen las funciones. Por ejemplo, se podría tener tres funciones  $f^{(1)}$ ,  $f^{(2)}$  y  $f^{(3)}$  conectadas en una cadena, para formar  $f(x) = f^{(3)}(f^{(2)}(f^{(1)}(x)))$ . Estas

estructuras de cadena son las estructuras más utilizadas de redes neuronales. En este caso,  $f^{(1)}$  se llama la primera capa de la red,  $f^{(2)}$  se llama la segunda capa, y así sucesivamente. La longitud total de la cadena determina la “profundidad” del modelo. La última capa es llamada capa de salida. Entre la capa de entrada y salida se encuentran las capas ocultas. La dimensión de las capas determina el ancho del modelo. También se puede pensar que cada capa consiste en muchas unidades que actúan en paralelo, cada uno representando una función de vector a escalar. La estructura de la arquitectura es similar a una NN clásica, pero como más capas, como se observa en la Fig. 2. 8.

### *2.9.1 Optimización de redes profundas*

Para sobrepasar la desaparición del gradiente y el sobreajuste en el proceso de aprendizaje, se aplican algunos algoritmos y técnicas de optimización en redes profundas como la deserción y el algoritmo de Adam. El algoritmo de Adam calcula las tasas de aprendizaje adaptativo individuales para diferentes parámetros a partir de las estimaciones de los primeros y segundos momentos de los gradientes. El nombre Adam se deriva de la estimación del Momento Adaptativo (del inglés, ADaptive Momentum) [15]. Adam almacena un promedio exponencial en descomposición de los gradientes cuadrados pasados  $v(t)$  y un promedio exponencial decreciente de los gradientes pasados  $m(t)$ , ambos similares a un momento, matemáticamente:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

donde  $\beta_1$  y  $\beta_2$  son hiperparámetros y  $g_t$  el gradiente. A continuación, se presenta el pseudocódigo presentado por Kingma y Lei (2015) (Kingma & Lei Ba, 2015):

---

**Algoritmo 1:**  $g_t^2$  indica el producto cruz cuadrado de  $g_t \odot g_t$ . Algunas configuraciones buenas para el testeo de problemas son:  $\alpha = 0.001$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  y  $\epsilon = 10^{-8}$ .

---

**Se requiere:**  $\alpha$ : Tamaño del paso.

**Se requiere:**  $\beta_1, \beta_2 \in [0,1)$ : Tasas de disminución exponencial para las estimaciones de momento.

**Se requiere:**  $f(\theta)$ : Función objetivo estocástica con parámetros  $\theta$ .

**Se requiere:**  $\theta_0$ : Vector del parámetro inicial.

$m_0 \leftarrow 0$  (Inicializar primer vector de momento)

$v_0 \leftarrow 0$  (Inicializar segundo vector de momento)

$t \leftarrow 0$  (Inicializar paso de tiempo)

**Mientras**  $\theta_t$  no converge, **hacer:**

$t \leftarrow t + 1$

$g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$  (Obtener gradientes)

$m_t \leftarrow \beta_1 * m_{t-1} + (1 - \beta_1) * g_t$  (Actualizar el primer momento estimado)

$v_t \leftarrow \beta_2 * v_{t-1} + (1 - \beta_2) * g_t^2$  (Actualizar el segundo momento estimado)

$\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$  (Calcular la corrección de las bias del primer momento estimado)

$\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$  (Calcular la corrección de las bias del segundo momento estimado)

$\theta_t \leftarrow \theta_{t-1} - \alpha * \frac{\hat{m}_t}{\sqrt{\hat{m}_t + \epsilon}}$  (Actualizar parámetros)

**termina mientras**

**regresa**  $\theta_t$  (Parámetros resultantes)

---

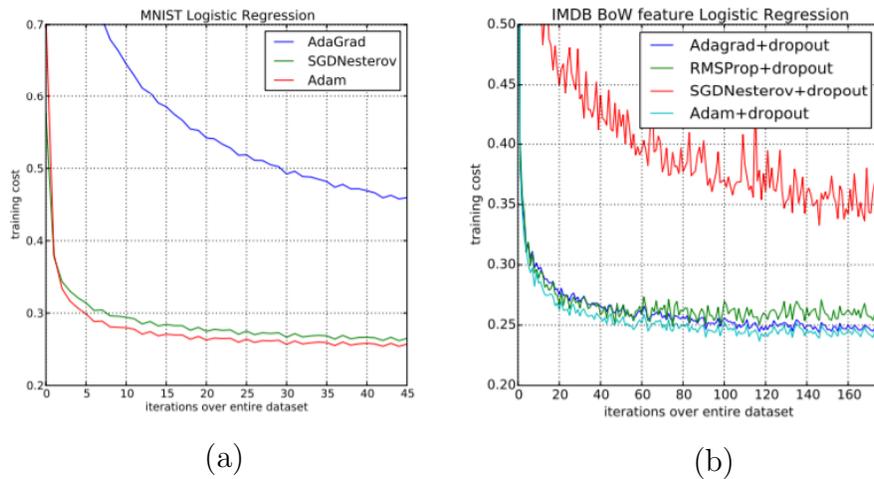


Fig. 2. 9. Resultados de regresión logística en entrenamientos con algoritmos de aprendizaje profundos (Kingma y Lei, 2015).

Como se observa en la Fig. 2. 9, publicado por Kingma y Lei (2015), la regresión logística del proceso de entrenamiento realizadas a bases de datos llamados por los autores como MNIST (a) y IMDB (b) aplicando diversos algoritmos de aprendizaje profundos, destacan los procesos ADAM y ADAM con deserción (dropout). Los métodos de regularización estocástica, como la deserción, permiten reducir el sobreentrenamiento. La idea clave de

la deserción es desaparecer unidades al azar y sus conexiones durante el entrenamiento. Básicamente, el proceso es conceptualizado en la Fig. 2. 10.

El mecanismo de deserción transforma una red neuronal profunda en un modelo reducido. Una red neuronal con  $n$  unidades se puede ver como la combinación de  $2^n$  posibles redes neuronales debilitadas. En tanto Adam es una combinación entre el Adagrad y el RMSprop y presenta mejor desempeño en diversas bases de datos.

El gradiente es una unidad de medida del error, y cada capa tiene un gradiente en particular. Este principio es la declaración matemática de los múltiples niveles de abstracción por lo que un modelo profundo se define. La combinación de la deserción con los algoritmos de aprendizaje profundo refuerza ciertos circuitos que constituye otra característica de las redes profundas, el apego por la forma de aprender en el cerebro biológico.

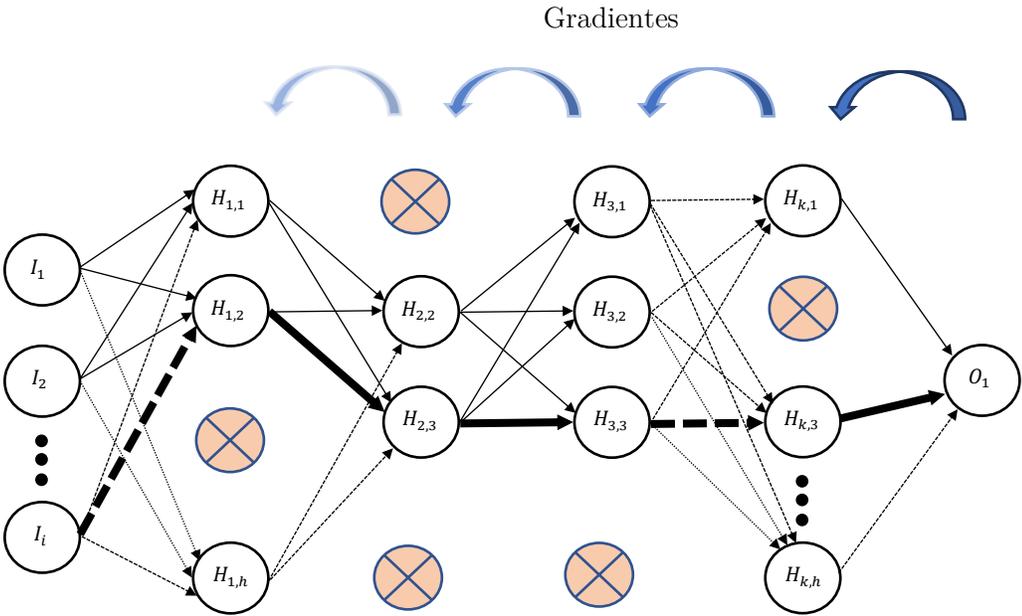


Fig. 2. 10. Desvanecimiento del gradiente y deserción en redes neuronales profundas.

## 2.10 Criterios estadísticos

Las ecuaciones estadísticas se utilizan con la finalidad de tener criterios que permitan comparar el desempeño de las arquitecturas de NN. A continuación, se describen algunas medidas estadísticas muy utilizadas para comparar grupos de datos generados con redes neuronales (Karamirad, Omid, Alimardani, Mousazadeth, & Heidari, 2013).

### 2.10.1 Coeficiente de correlación

El coeficiente de correlación de Pearson  $R$  es una medida de la relación lineal entre dos variables aleatorias cuantitativas. Este coeficiente es un índice que no depende de las magnitudes de las variables, siempre y estas sean cuantitativas. Matemáticamente se puede relacionar los datos experimentales y simulados con la fórmula:

$$R = \frac{\sum_{i=1}^N (y_i - y_m)(s_i - s_m)}{\sqrt{\sum_{i=1}^N (y_i - y_m)^2 \sum_{i=1}^N (s_i - s_m)^2}} \quad \text{Ec. 2. 11}$$

donde:  $N$  es la cantidad total de datos.  $y_i$  es el valor deseado de la salida  $i$ -ésima,  $y_m$  es la media de las salidas esperadas,  $s_i$  es la salida  $i$ -ésima entregada por la NN,  $s_m$  es la media de las salidas entregadas por la NN.

### 2.10.2 Error Absoluto Medio Porcentual

El MAPE (Mean Absolute Percent Error, por sus siglas en inglés) mide la magnitud del error en términos porcentuales. Este es calculado como el promedio del error porcentual sin signo. Matemáticamente:

$$MAPE = \frac{1}{N} \sum_i^N \left| \frac{y_i - s_i}{y_i} \right| \quad \text{Ec. 2. 12}$$

Para valores  $y_i$  pequeños, el MAPE toma frecuentemente resultados extremos. Como este número es un porcentaje, puede ser interpretado más fácilmente que otros métodos estadísticos.

### 2.10.3 Error Cuadrático Medio

El error cuadrático medio (MSE o mean square error en inglés) es una forma de evaluar la diferencia entre un estimador o predictor y el valor real de la cantidad que se quiere calcular. El MSE mide el promedio del cuadrado del error, matemáticamente es:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - s_i)^2 \quad \text{Ec. 2. 13}$$

---

## Capítulo 3

### Metodología

---

En esta sección se explicará la metodología para lograr el objetivo de este estudio. En principio, es necesario tener una base de datos con las muestras de potencia eléctrica de una planta fotovoltaica, adecuar la información para poder procesarla con los modelos de inteligencia artificial y seguidamente ajustar los parámetros de cada método de pronóstico para obtener la mejor predicción. El desempeño de cada modelo será determinado por las ecuaciones de criterio mencionadas en el capítulo anterior, los cuales son indicadores del error que existe entre el valor pronosticado y los valores pasados que se ingresen. Debido a que los modelos tienen variaciones en cuanto a sus parámetros y a su ajuste, se presenta un método de comparación con los atributos particulares. Los detalles del diseño se explican en los apartados siguientes.

En la Fig. 3. 1, se observa el diagrama global de los pasos que se utilizarán en la metodología de este trabajo. Es posible distinguir tres grandes bloques de procesos en el diagrama: El primero se caracteriza por la instrumentación de la planta fotovoltaica para la adquisición de datos, la segunda toma en cuenta los procesos de transformación y procesamiento de la información recolectada y finalmente la aplicación de los métodos de pronósticos.

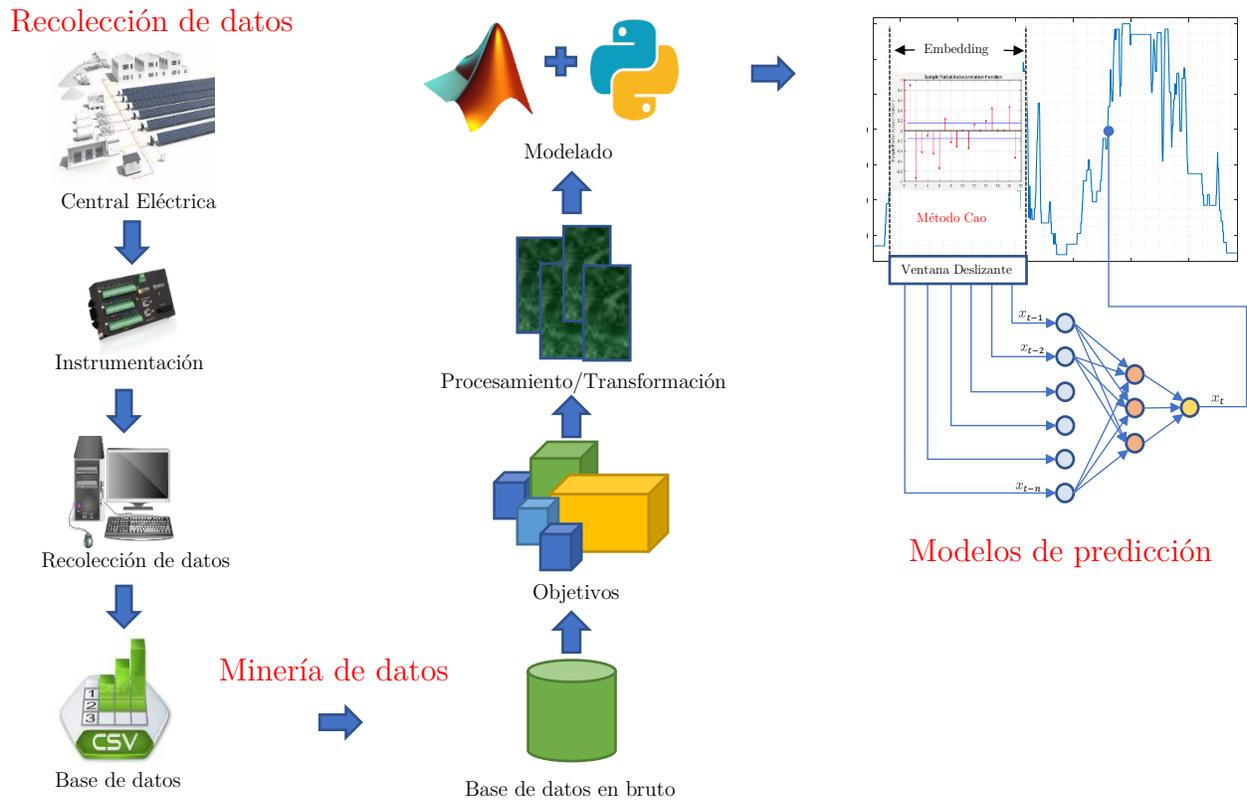


Fig. 3. 1. Diagrama de la metodología. Se compone de la recolección, minería de datos y aplicación de modelos de predicción.

### 3.1 Recolección de datos

Las centrales eléctricas fotovoltaicas o plantas fotovoltaicas del presente estudio se encuentran en el noroeste de México. Durante el año 2015 se recolectaron muestras cada minuto, dando un total de 525, 600 pares de datos. Se hicieron recolecciones de dos plantas, teniendo un total de 1,051,200 datos. El proceso de medición de datos se encuentra estrictamente normado por instituciones como el SENER, CENACE y CFE.

Se agradece a la M.I Cielo Bautista por haber proporcionado los datos necesarios para llevar a cabo este estudio. El primer bloque del diagrama se encuentra implícitamente realizado por el trabajo de recolección del CENACE, donde la instrumentación fue llevada a cabo mediante las normas CFE G0000-48. En la Fig. 3. 2 se encuentran las gráficas de las series de tiempo para el mes de enero de la planta CVERDESOLAR y DREW.

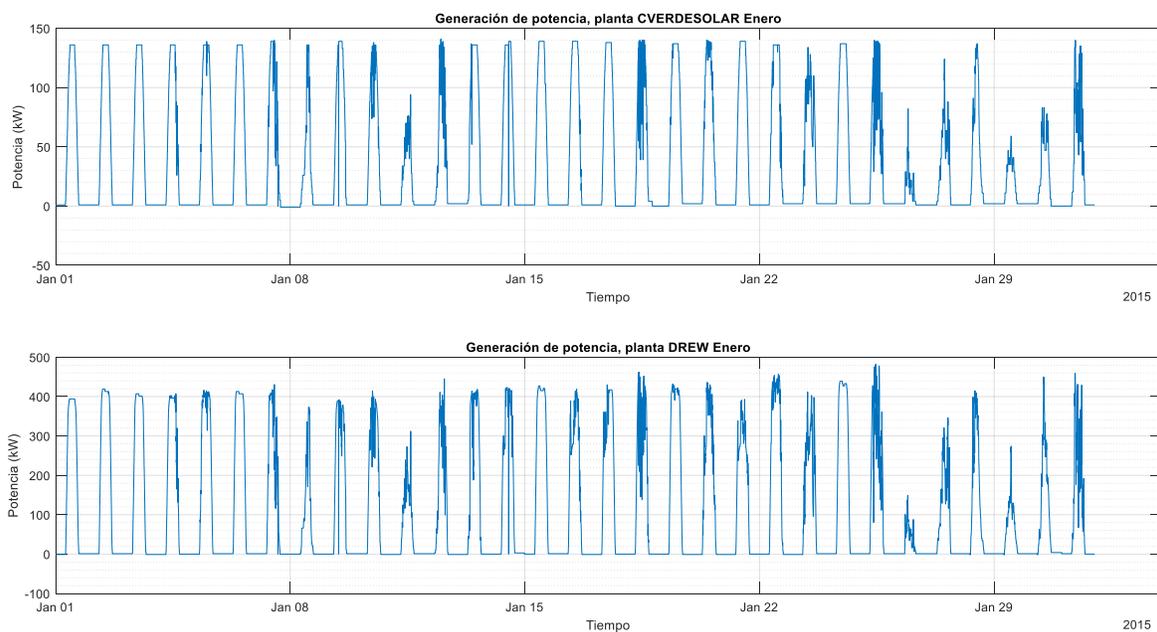
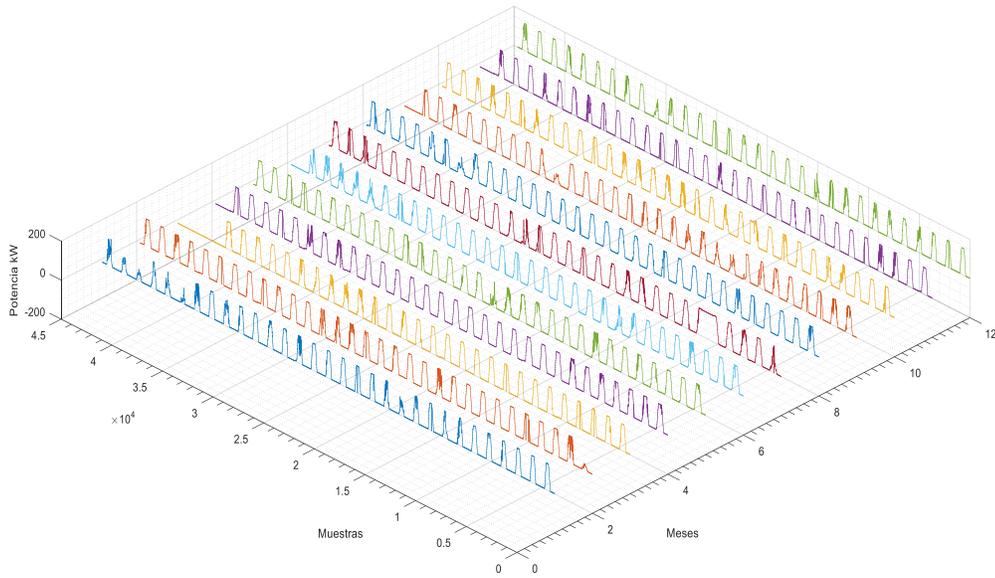


Fig. 3. 2. Gráficas de potencia eléctrica recolectadas en las plantas DREW y CVERDESOLAR en el mes de enero del 2015 (CENACE).

(a) CVERDE SOLAR



(b) DREW

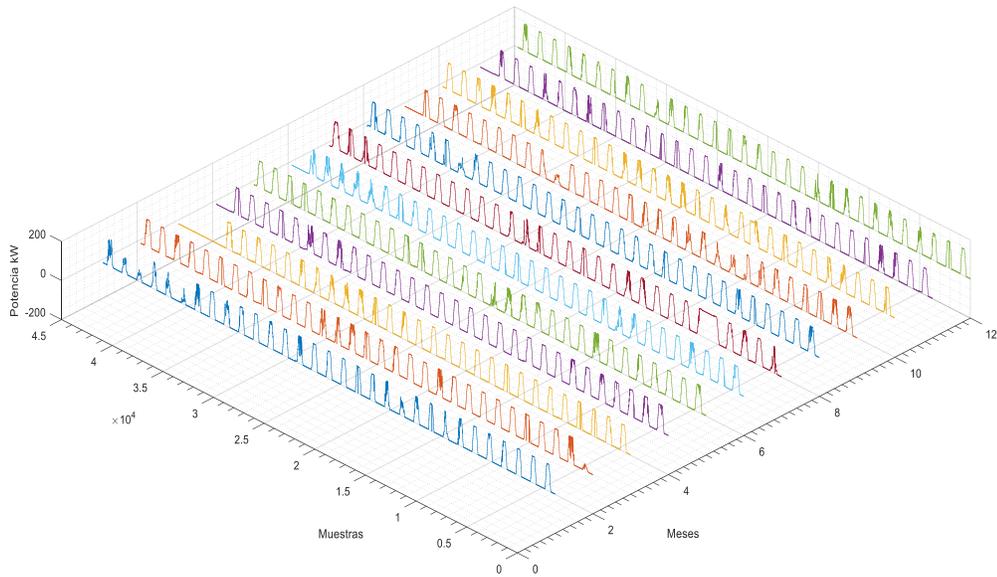


Fig. 3. 3. Gráficas 3D de potencia eléctrica recolectadas en las plantas CVERDESOLAR y DREW en el año 2015 (CENACE).

Con la finalidad de observar el comportamiento de la potencia eléctrica en todo el año de la base de datos CVERDESOLAR, en la Fig. 3.3 (a) se presenta una vista 3D de las muestras de datos separadas por meses. Se observa que cada campana corresponde a un día en particular y no se tiene un comportamiento lineal. La Fig. 3.3 (b) corresponde a la segunda base de datos DREW.

### 3.2 Minería de datos

Dada la base de datos proporcionada, es necesario aplicar un tratamiento a la información para que se adecue a los modelos de predicción. Algunas herramientas de la metodología de minería de datos son utilizadas para descubrir *conocimientos entre los datos*. La minería de datos puede ser vista como el resultado de la evolución natural de la información tecnológica. Muchas personas tratan la minería de datos como un sinónimo de otro término popularmente utilizado, el “descubrimiento de conocimiento a partir de datos”, o KDD por sus siglas en inglés, mientras que otros ven la minería de datos simplemente como un paso esencial en el proceso de descubrimiento de conocimiento (Han, Kamber, & Pei, 2012).

Aplicando el proceso KDD a la base de datos, los pasos de la metodología de minería serían los siguientes:

1. *Limpieza de datos*. Sirve para remover ruido y datos inconsistentes. La base de datos ha pasado por este tratamiento, por lo cual no se observan muestras fuera de lugar, y no hay interrupciones por toda la serie.

2. *Integración de datos.* Los datos necesitan ser unidos para formar una secuencia completa del año, pues la información se encuentra separada por meses, y debido a que el proceso de aprendizaje de los modelos debe ser en su totalidad, este proceso debe hacerse con cuidado.
3. *Selección de datos.* En primera instancia, se deben seleccionar los datos de potencia eléctrica durante el tiempo de luz solar, por lo cual se debe realizar un algoritmo que filtre los datos medidos de noche. En segundo lugar, se deben seleccionar datos para la validación y entrenamiento de los modelos de predicción.
4. *Transformación de datos.* La base de datos proporcionada se encuentra en periodos de muestreo de 1 min. Debido a que se debe realizar la predicción al día siguiente, los datos deben ser promediados por horas para que la predicción se realice con 24 pasos de anticipación, y/o promediada por días, para aplicar la predicción con un paso de anticipación. La transformación de datos también implica la reestructuración de la serie de tiempo de una dimensión a un arreglo  $N \times M$ , donde  $M$  representa las columnas para la dimensión del embedding y el paso siguiente (o pasos siguientes) y  $N$  la dimensión de total de muestras dado el tamaño de ventana para predicción.
5. *Minería de datos.* En este paso, se extrae la esencia de los fenómenos que se encuentran implícitos en la serie de tiempo y, que servirán para la predicción de los valores futuros.
6. *Evaluación de patrones.* En este punto se utiliza la base de datos de validación para evaluar con criterios estadísticos el desempeño de los modelos.
7. *Presentación del conocimiento.* Usualmente se presentan los datos en gráficas o esquemas que cualquier usuario pueda entender con facilidad.

La minería de datos tiene una conexión inherente con las estadísticas. Las estadísticas estudian la recopilación, análisis, interpretación o explicación y presentación de datos. El presente estudio elaborará algunas estadísticas descriptivas como la media, moda y desviación estándar para los datos, separados por meses. También es importante tener una medición de la desviación estándar cuando se realicen los procesos de transformación de datos.

El proceso de transformación sigue el algoritmo de la Fig. 3. 4. En este algoritmo se considera la hora de amanecer y de anochecer, que debe ser calculada con el huso horario, latitud y longitud de la localización de la planta eléctrica fotovoltaica. Al considerar las horas de luz, se filtra las horas de noche, las cuales no son necesarias. Con la finalidad de promediar los datos por horas, una matriz de tres dimensiones debe ser inicializada, para colocar los promedios por horas, correspondientes a la hora en particular (primera dimensión), del día en cuestión (segunda dimensión) del mes a tratar (tercera dimensión).

Cada mes tiene su número de días, por lo cual el algoritmo debe ser capaz de conocer el número de días en el mes fraccionando el formato de fechas (ejemplo: 26/04/1998) y rastreando el último dato. Finalmente se hace un barrido de todos los datos para hacer los promedios siguiendo el diagrama de flujo de la Fig. 3. 4.

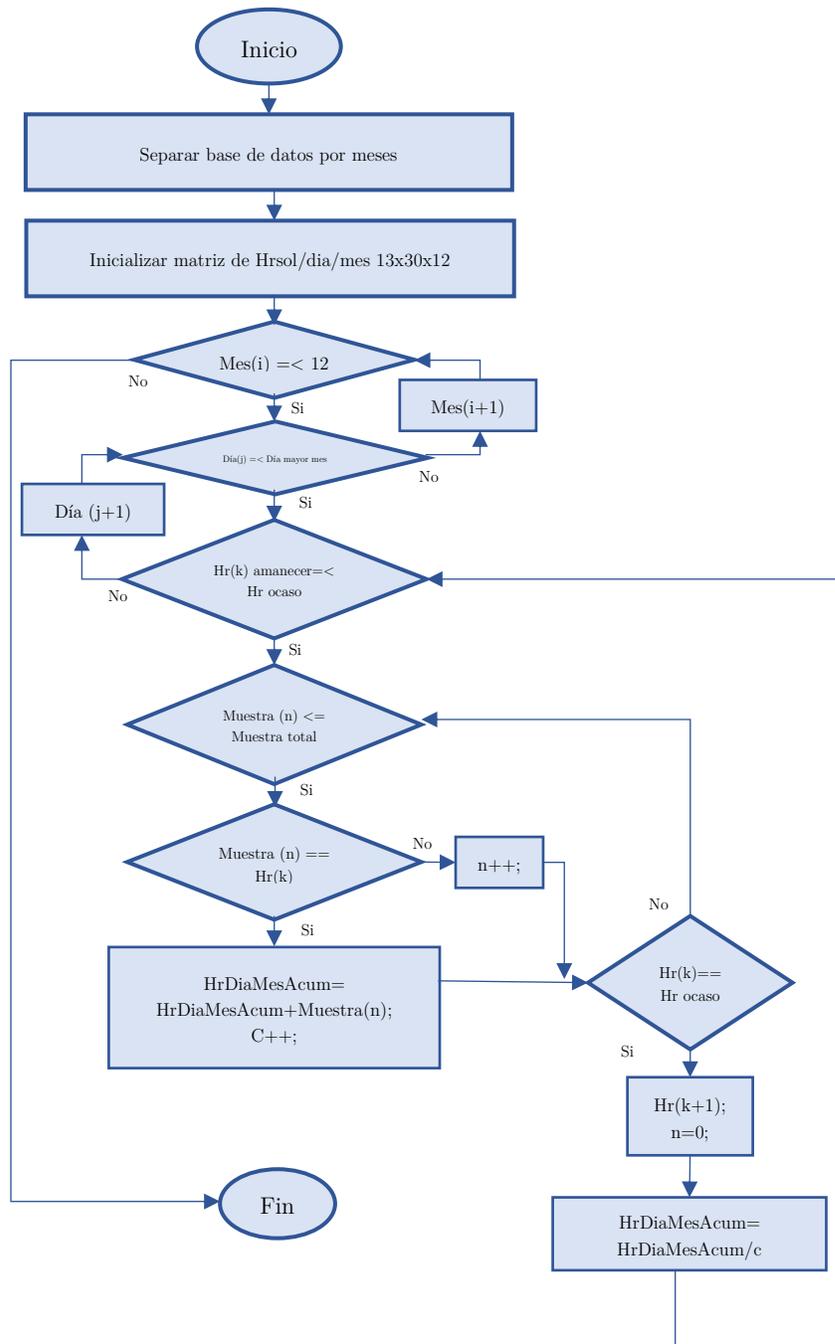


Fig. 3. 4. Diagrama de flujo para la extracción de la potencia en horas de Sol.

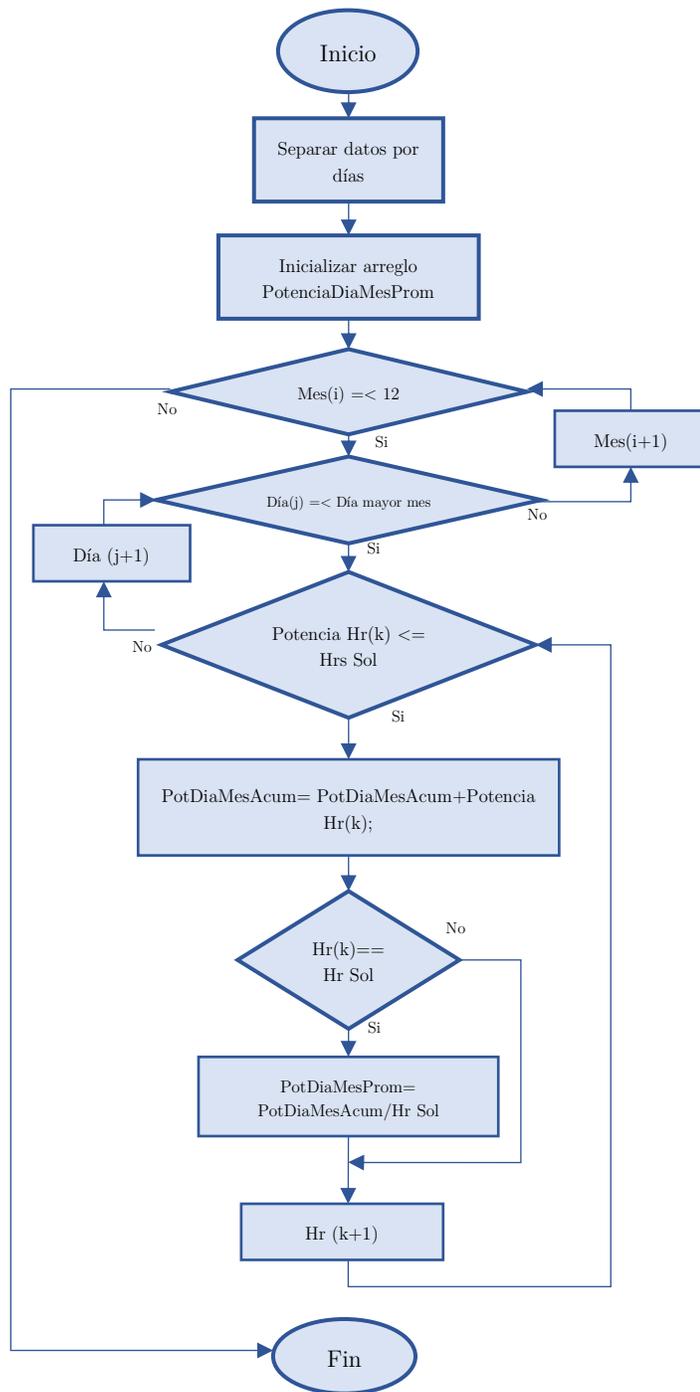


Fig. 3. 5. Diagrama de flujo para el promedio por horas de la potencia eléctrica fotovoltaica.

Luego de realizar los promedios, se procede a realizar los cálculos para obtener la media por día, siguiendo un algoritmo similar al anterior. La Fig. 3. 5 muestra el diagrama de flujo para convertir la serie de tiempo anterior en una nueva con muestras horarias. Con este proceso de transformación, se obtiene la base de datos lista para separar en datos de validación y entrenamiento.

### 3.3 Modelos de predicción

Ya con la serie de tiempo de los datos de potencia eléctrica, es posible aplicar los modelos de predicción seleccionados para este estudio. Los modelos en cuestión son las redes neuronas clásicas NN, las redes neuronales profundas DNN y el ARMA. Para aplicar estos

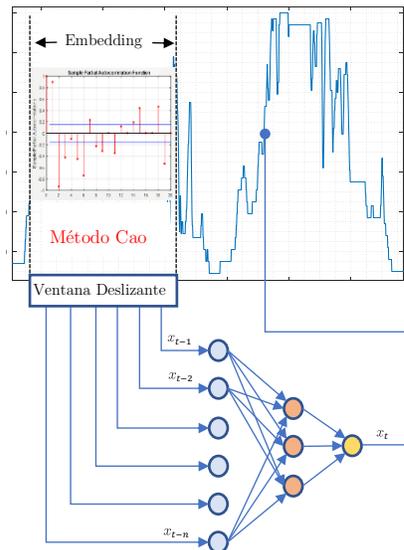


Fig. 3. 6. Procedimiento para el entrenamiento de los modelos. La ventana de dimensión de valores pasados se determina con el Método de Cao.

modelos, es necesario determinar el tamaño del embedding o ventana de predicción (datos pasados) para predecir un valor futuro.

Existe en la literatura métodos heurísticos para determinar el ancho de la ventana de embedding, el cual consiste en incrementar el número de datos pasados hasta obtener la mejor predicción. Con la finalidad de establecer un método determinístico para proponer la ventana, el Método de Cao es aplicado. El método fue descrito en el marco teórico, y hasta este punto de la metodología es aplicado usando funciones de Matlab.

Como se observa en la Fig. 3. 6, los datos pasados de la ventana, servirán como entradas de los modelos de predicción. La salida del modelo representa el dato futuro, que en la serie de tiempo corresponde al valor de potencia eléctrica del paso siguiente. Se describe el embedding como el número de entradas en los modelos, y a continuación, se describirá el proceso de ajuste de los modelos para obtener el mejor desempeño.

### **3.4 Entrenamiento de modelos inteligentes**

Los modelos de inteligencia artificial requieren un proceso heurístico para ajustar los parámetros de red. Este ajuste permite alcanzar la mejor predicción posible del modelo. Dado que las neuronas en la capa de entrada son establecidas por el método de Cao, las neuronas en la capa oculta, las funciones de activación, la profundidad del modelo y el algoritmo de aprendizaje son los parámetros que pueden cambiar.

El diagrama del proceso de entrenamiento es ilustrado en la Fig. 3. 7. Se encuentra en el esquema cuatro secciones principales: Base de datos, Modelos IA, Entrenamiento y

Criterio Estadístico del error. Las secciones se relacionan con las demás, mediante un flujo direccionado. Los detalles se presentan a continuación.

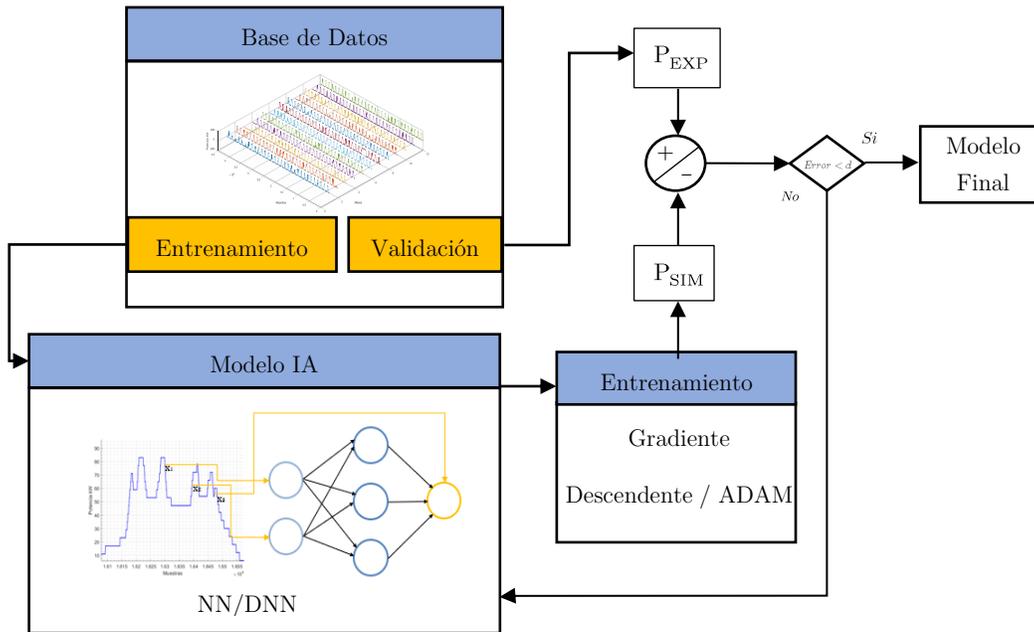


Fig. 3. 7. Proceso de entrenamiento para los modelos de inteligencia artificial.

### 3.4.1 Base de datos

El proceso de formación de la base de datos ha sido comentado con anterioridad. Para el proceso de ajuste de los modelos, se divide la base de datos en: 80 % para el entrenamiento y 20 % para propósitos de validación. Esto se traduce en que la serie de tiempo de potencia eléctrica será dividida en dos partes. Los datos de entrenamiento son pasados a los modelos IA, que representan al siguiente bloque.

### 3.4.2 Modelos IA

Para obtener la arquitectura final se necesita variar los parámetros de los modelos IA. El modelo NN clásico consta de una capa oculta, como fue considerado en el marco teórico. Las funciones de transferencia se organizan por pares, que representan a la capa oculta-salida. Debido a que la salida de la red neuronal es continua, la función de activación de la neurona de salida se establece como purelin. Entonces, se tiene que el ajuste de las funciones de transferencia viene dado por los pares logsig-purelin y tansig-purelin. Finalmente, el número de neuronas se incrementa en cuanto el proceso iterativo lo requiera. Para dejar en claro, la arquitectura se presenta como  $I-H-1$ , donde I es el número de neuronas de entrada determinado por el método de Cao, H representa al número de neuronas en la capa oculta y 1, que representa al paso siguiente en el tiempo.

El modelo DNN contiene los mismos parámetros que el modelo NN. La diferencia radica en que se incrementan las capas ocultas de la arquitectura. Para denominar la arquitectura profunda, se propone lo siguiente:  $I-H_1-...-H_m-1$ . Donde la H se refiere al número de neuronas en la capa y el subíndice a la profundidad. Con lo que respecta a las funciones de transferencia, se puede considerar que cada capa tiene una ecuación diferente. Esto formaría cadenas de funciones como tansig-tansig-logsig-purelin, para una DNN con tres capas ocultas. El purelin se mantiene igualmente para la capa de salida.

Estas arquitecturas serán transferidas al siguiente bloque, el referido al entrenamiento de los modelos.

### *3.4.3 Entrenamiento*

El entrenamiento de las arquitecturas permite ajustar los pesos y sesgos de las redes neuronales para minimizar el error entre los datos pronosticados y los experimentales. Existen diversas librerías que tienen las herramientas computacionales referidas a Inteligencia Artificial, Machine Learning y Deep Learning. Para el entrenamiento de las redes neuronales, se propone utilizar la API de redes neuronales de Keras. Esta librería es escrita en Python y capaz de ejecutarse sobre TensorFlow, CNTK o Theano.

El desarrollo del código consistió en un reto, dada la reciente emergencia de los algoritmos de aprendizaje profundo. Con la finalidad de obtener los datos estadísticos, la librería de Scikit-learn fue instalada en la computadora. El sistema operativo para la implementación de todas estas librerías es Ubuntu 14.1, pues es compatible y desarrollado para ese entorno. Mediante las herramientas computacionales mencionadas anteriormente, se implementó el código de entrenamiento para los modelos de redes neuronales clásicos NN y los profundos DNN.

Para la comparación de los modelos, el aprendizaje con gradiente descendente es aplicado al modelo NN y el algoritmo de Adam para las DNN. Los algoritmos marcan la diferencia fundamental entre los Modelos IA. Los modelos profundos carecían de uso y popularidad debido a los problemas de sobreentrenamiento y desvanecimiento del gradiente. Dentro del código se implementó la técnica de Dropout para prevenir la especialización de la red.

Después del entrenamiento de los modelos con Keras, se realiza la evaluación del desempeño con los criterios estadísticos, marcando el inicio del próximo bloque.

#### *3.4.4 Error estadístico*

El entrenamiento del modelo seleccionado ajusta los pesos para tener el mejor desempeño con la arquitectura definida en el segundo bloque. Sin embargo, no es posible definir si el cambio de una función de transferencia, el incremento de neuronas o el incremento de la profundidad mejorará las predicciones. Es debido a esto, que los valores pronosticados de potencia eléctrica fotovoltaica son comparados con los datos de validación experimentales provenientes de la etapa de base de datos. Si el error es mayor a un umbral definido por el usuario, el flujo y dirección del esquema se dirige a la etapa de Modelos IA, donde la arquitectura es redefinida. En primera instancia, se cambian los valores de las funciones de transferencia y se repite el proceso. Seguidamente, se vuelve a entrenar. Si el error sigue siendo excesivo, se repite el proceso, modificando los números de neuronas en las capas ocultas. Finalmente, para los modelos DNN, se incrementarán las capas ocultas.

---

## Capítulo 4

### 4. Resultados

---

En este capítulo se expondrán los resultados obtenidos de la metodología. Con el método de Cao se calculó los anchos de ventana de predicción, que sirvió para determinar la dimensión de entrada de los modelos de predicción. Seguidamente se aplicaron los diversos modelos para obtener las mejores predicciones.

#### 4.1 Procesamiento de la base de datos

La base de datos se encontraba en muestras cada 5 minutos permitiendo el pronóstico a corto plazo. Como el objetivo de la tesis es la predicción al día siguiente, la base de datos se estructura en horas, con la finalidad de observar el comportamiento, y luego a días, como se muestran en las Fig. 4. 1 y 4. 2.

Con finalidad de obtener los datos estadísticos de este procedimiento, en la Fig.4. 3 se presenta la desviación estándar de los datos promediados.

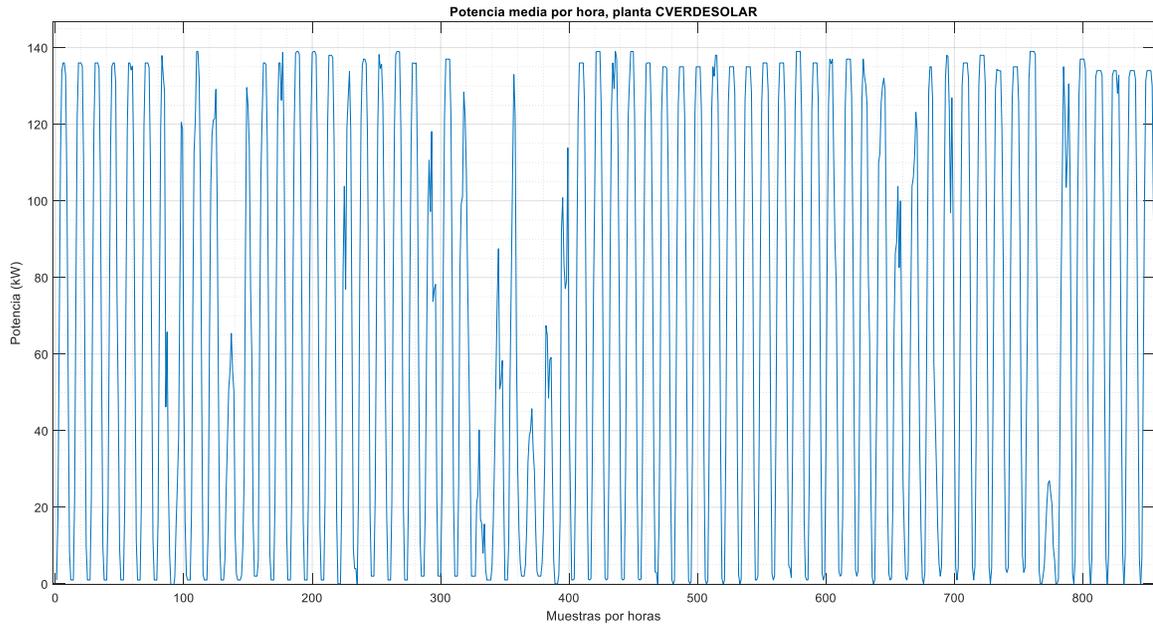


Fig. 4. 1. Base de datos procesada por horas.

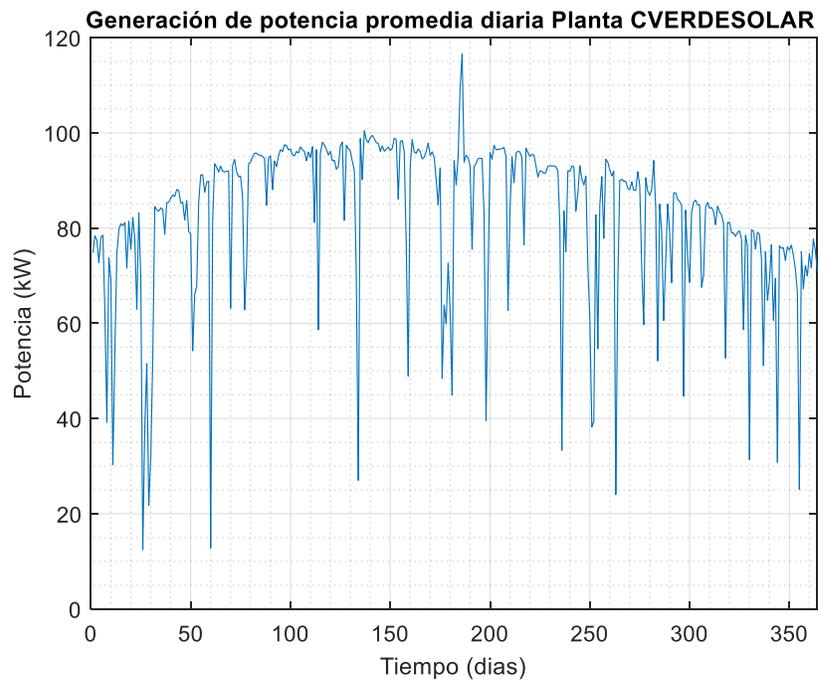


Fig. 4. 2. Base de datos procesada por días.

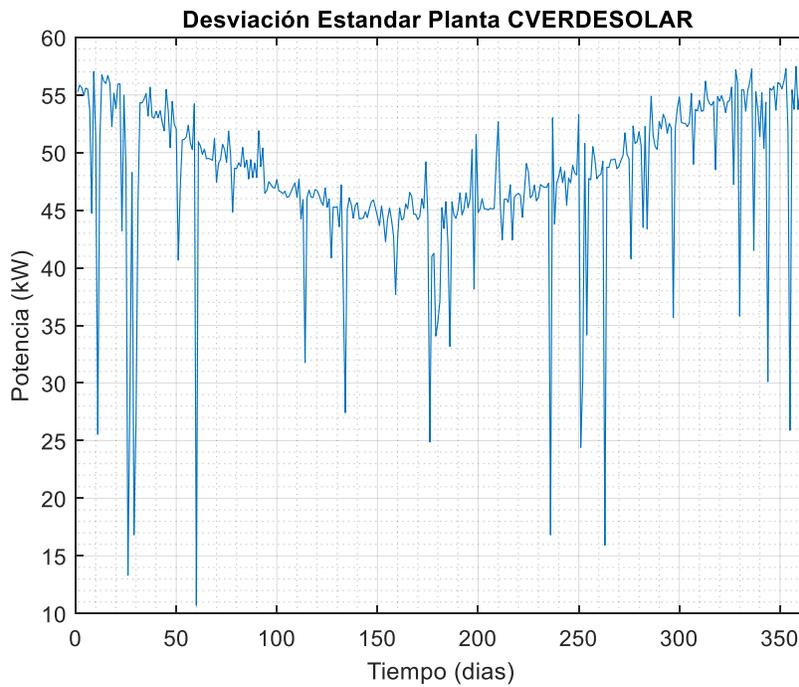


Fig. 4. 3. Desviación estándar de los promedios diarios de la base de datos.

## 4.2 Dimensión de embebimiento

Mediante un script elaborado en Matlab, se procedió a calcular la función de autocorrelación parcial de muestra que utiliza como parámetro de entrada el Método de Cao. En la Fig. 4. 5 se puede observar que con un retraso de cuatro se tiene una autocorrelación para el Método de cao adecuada. Tomando en cuenta que la dimensión de embebimiento se determina con los promedios de distancia de falsos vecinos cercanos a la unidad, las dimensiones pueden tener de 5 a 9 entradas aceptables para los modelos de predicción.

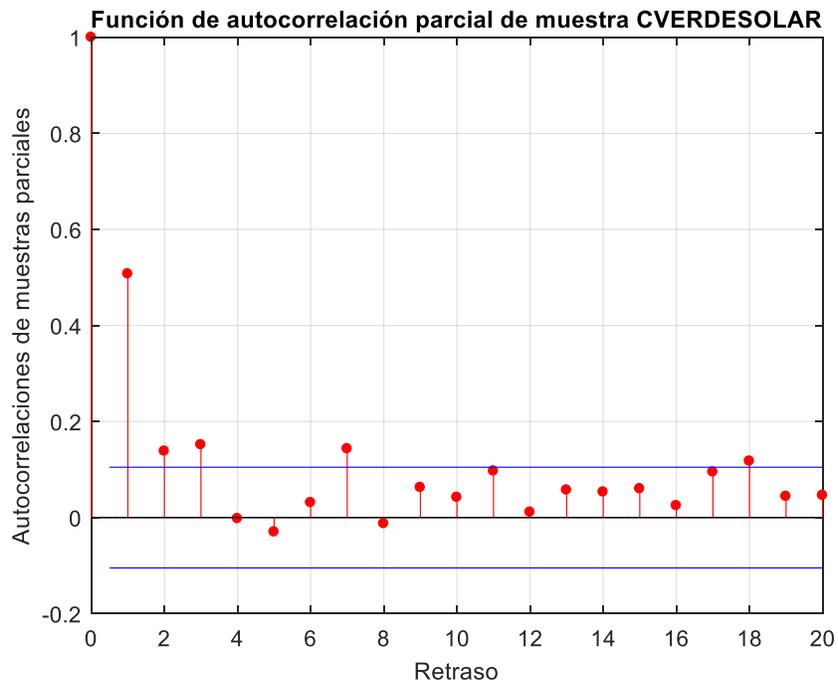


Fig. 4. 4. Cálculo de autocorrelaciones para el Método de Cao.

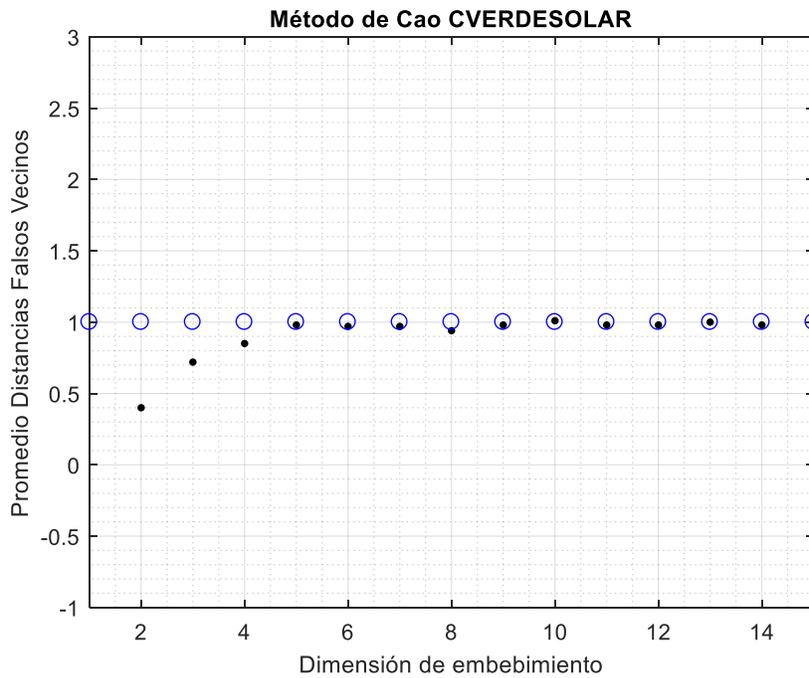


Fig. 4. 5. Determinación de la dimensión de embebimiento por Cao.

### 4.3 Predicción con ARMA

Para ajustar los valores de los órdenes polinomiales  $na$  y  $nc$  del ARMA fue utilizado la función de aproximación de espacio de estados del Matlab con el 80% de la base de datos.

El modelo de tiempo discreto  $A(z)y(t) = C(z)e(t)$  es el siguiente:

$$A(z) = 1 - 0.5527 z^{-1} - 0.9449 z^{-2} + 0.4973 z^{-3}$$

$$C(z) = 1 - 0.09492 z^{-1} - 0.9629 z^{-2} + 0.132 z^{-3}$$

En la Fig. 4. 6 se observa los datos de potencia  $P_{train}$  obtenidos del ajuste del modelo ARMA y la predicción  $P_{test}$  utilizando un día de anticipación. Para obtener las predicciones se hizo un barrido de la base de datos de prueba con la ventana de 7 obtenida con Cao.

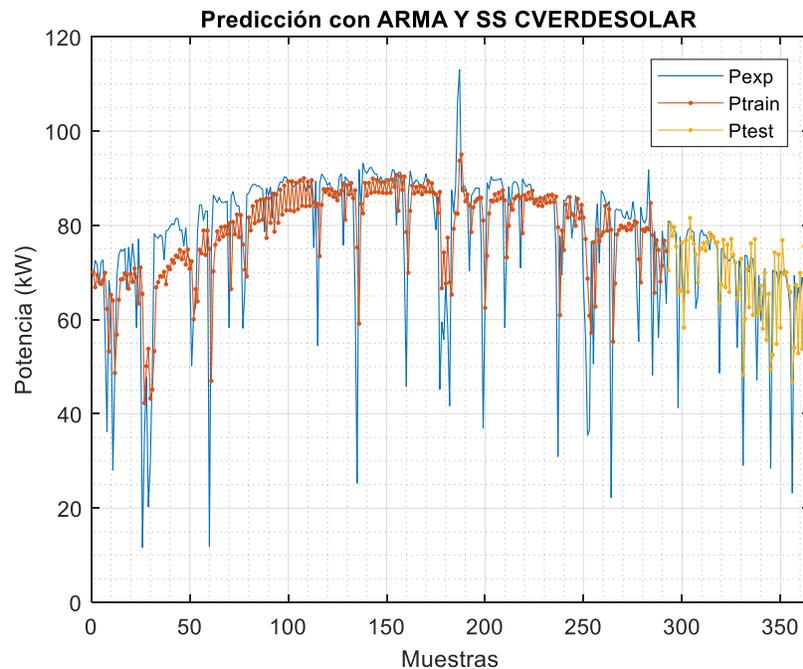


Fig. 4. 6. Base de datos de entrenamiento y de prueba con ARMA.

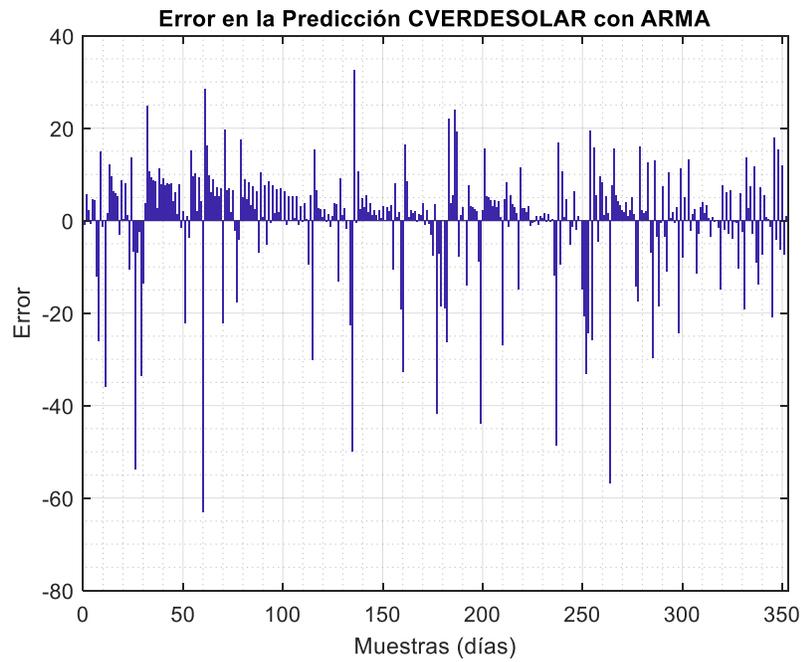


Fig. 4. 7. Diferencia entre la base de datos y la predicción con ARMA.

En la Fig. 4. 7 se observa la diferencia entre los valores de la base de datos y los datos calculados con el arma. Se observa que el error es constante durante el entrenamiento y la de prueba.

Los errores obtenidos en el proceso de ajuste se encuentran en la Tabla 4. 1.

Tabla 4. 1. Criterios estadísticos del ajuste con ARMA.

Modelo	Criterios estadísticos								
	Entrenamiento			Prueba			Total		
	RMSE	MAPE	R	RMSE	MAPE	R	RMSE	MAPE	R
ARMA	12.8351	16.7579	0.5612	8.8117	11.8472	0.6552	12.1376	15.7757	0.6132

## 4.4 Predicción con Redes neuronales

### 4.4.1 Selección de la función de transferencia

Con la finalidad de obtener la función de transferencia que mejor se ajuste a los datos, se realizaron entrenamientos con una neurona en la capa oculta y se incrementó la ventana de predicción tomando en cuenta para cada uno los pares de funciones de transferencia logsig-purelin y tansig-purelin. Para cada entrenamiento, se aplicaron mil épocas con 100 interacciones. En la Tabla 4. 2 se puede observar que, para cada ventana de predicción, el par tansig-purelin es el que tiene los menores errores y las correlaciones más altas. De igual manera, se presenta en la tabla los valores mínimos y máximos y su diferencia, notando que, en los datos de prueba, el criterio MAPE es el error más significativo para cada ajuste.

Tabla 4. 2. Criterios estadísticos del ajuste con NN.

Arquitectura			Función Activación	Criterios estadísticos								
				Entrenamiento			Prueba			Total		
I	H	O		RMSE	MAPE	R	RMSE	MAPE	R	RMSE	MAPE	R
1	1	1	logsig-purelin	13.1514	17.4893	0.5275	13.4103	17.4544	0.1162	13.2028	17.4824	0.517
1	1	1	tansig-purelin	13.0567	17.3386	0.5372	13.0767	16.764	0.1327	13.0606	17.2256	0.5309
2	1	1	logsig-purelin	13.1508	17.4677	0.5295	13.223	17.1513	0.1157	13.1649	17.406	0.5239
2	1	1	tansig-purelin	13.1119	17.4083	0.5335	13.0887	16.842	0.121	13.1074	17.2979	0.5292
3	1	1	logsig-purelin	13.0818	17.3552	0.5384	13.1569	16.9733	0.1158	13.0964	17.2814	0.5341
3	1	1	tansig-purelin	13.0732	17.3007	0.5393	13.1401	16.9382	0.1172	13.0861	17.2306	0.5356
4	1	1	logsig-purelin	13.092	17.3553	0.5384	13.3123	17.4156	0.1042	13.1345	17.3668	0.5338
4	1	1	tansig-purelin	13.083	17.3268	0.5393	13.2671	17.3092	0.1053	13.1185	17.3234	0.5353
5	1	1	logsig-purelin	13.0949	17.4562	0.5399	13.4472	17.7874	0.0911	13.1625	17.5191	0.5339
5	1	1	tansig-purelin	13.0893	17.4996	0.5408	13.3358	17.5405	0.0919	13.1364	17.5073	0.5353
			Mínimo	13.0567	17.3007	0.5275	13.0767	16.764	0.0911	13.0606	17.2256	0.517
			Máximo	13.1514	17.4996	0.5408	13.4472	17.7874	0.1327	13.2028	17.5191	0.5356
			Diferencia	0.7%	1.1%	2.5%	2.8%	5.8%	31.3%	1.1%	1.7%	3.5%

#### 4.4.2 Determinación de la ventana de predicción

Ya definida la función de transferencia para la capa oculta, el siguiente paso consiste en determinar si la ventana de dimensionamiento obtenida con el método de Cao es consistente con el método heurístico popular en redes neuronales para determinar el número de neuronas en la capa de entrada. En la Tabla 4. 3 se muestran los resultados del incremento de neuronas de entrada (o ventana de predicción) con los errores y correlaciones obtenidas en los ajustes de los pesos y bias.

Tabla 4. 3. Criterios estadísticos para la selección de ventana.

Arquitectura			Criterios estadísticos								
			Entrenamiento			Prueba			Total		
I	H	O	RMSE	MAPE	R	RMSE	MAPE	R	RMSE	MAPE	R
1	1	1	13.0567	17.3386	0.5372	13.0767	16.764	0.1327	13.0606	17.2256	0.5309
2	1	1	13.1119	17.4083	0.5335	13.0887	16.842	0.121	13.1074	17.2979	0.5292
3	1	1	13.0732	17.3007	0.5393	13.1401	16.9382	0.1172	13.0861	17.2306	0.5356
4	1	1	13.083	17.3268	0.5393	13.2671	17.3092	0.1053	13.1185	17.3234	0.5353
5	1	1	13.0893	17.4996	0.5408	13.3358	17.5405	0.0919	13.1364	17.5073	0.5353
6	1	1	13.1072	17.4385	0.5406	12.7221	16.6047	0.1339	13.0356	17.2818	0.5402
7	1	1	<b>13.0153</b>	<b>17.3117</b>	<b>0.5477</b>	<b>12.4408</b>	<b>16.5123</b>	<b>0.2029</b>	<b>12.9102</b>	<b>17.1628</b>	<b>0.5539</b>
8	1	1	12.9048	17.1699	0.5435	12.4772	16.5823	0.206	12.827	17.0615	0.5512
9	1	1	12.9097	17.1204	0.544	12.5055	16.6161	0.2097	12.8369	17.3293	0.552
10	1	1	12.9984	17.5665	0.532	12.8912	16.5912	0.1945	12.8923	17.3299	0.555
Mínimo			12.9048	17.1204	0.532	12.4408	16.5123	0.0919	12.827	17.0615	0.5292
Máximo			13.1119	17.5665	0.5477	13.3358	17.5405	0.2097	13.1364	17.5073	0.555
Diferencia			1.6%	2.5%	2.9%	6.7%	5.9%	56.2%	2.4%	2.5%	4.6%

Debido a que el incremento del número de neuronas corresponde a cambios muy pequeños en los errores y correlaciones, se determinaron los mínimos y máximos de cada ajuste de red para decidir cuál es el número de entrada que tiene un impacto mayor. Para sorpresa, el método de Cao determina con bastante certidumbre que con una ventana de 7

elementos es posible tener buenos resultados. En la Tabla 4. 3 se observa que la correlación en el proceso de prueba tiene un impacto mayor en los criterios estadísticos, mientras que en los demás los cambios tienen una presencia muy despreciable.

En la Fig. 4. 8 se observa como los datos ajustaron para los datos de entrenamiento y de prueba. Los errores entre la base de datos real y las predicciones con redes se presentan en la figura. El comportamiento es parecido a la predicción con ARMA, pero con criterios estadísticos más bajos.

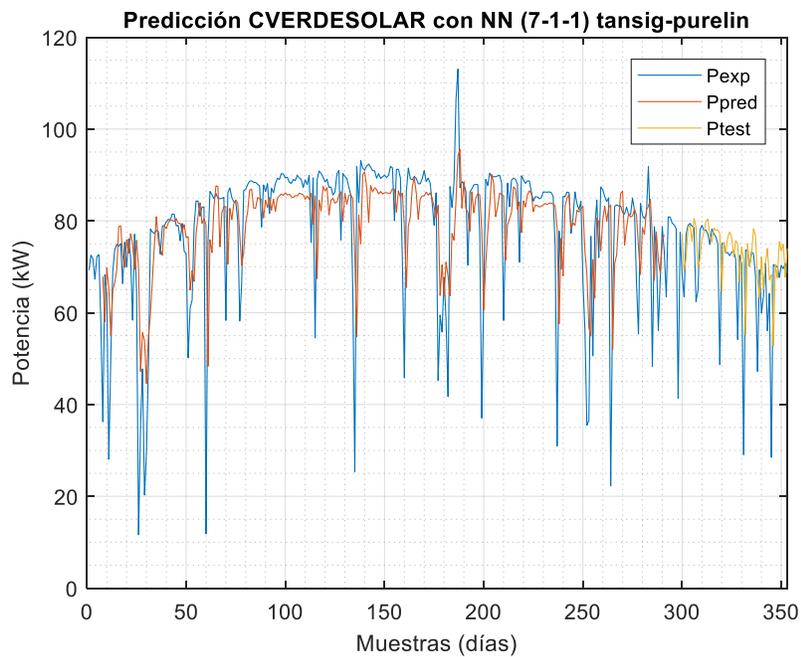


Fig. 4. 8. Comportamiento de las predicciones con la ventana de predicción.

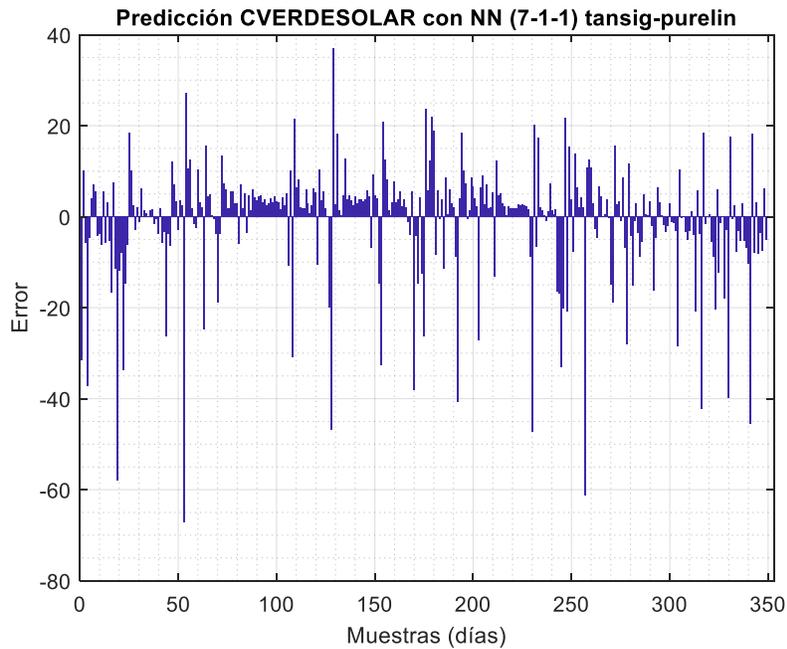


Fig. 4. 9. Diferencia entre la base de datos y la predicción con NN.

#### 4.4.3 Selección de la función de transferencia

Finalmente, con la ventana de predicción y la función de transferencia ya definidas, el siguiente paso consiste en ajustar el modelo de red neuronal en orden para obtener la arquitectura que mejor desempeño presente. Lo anterior corresponde al número de neuronas en la capa oculta y se incrementaron las unidades hasta no obtener mejores resultados. La consideración sigue siendo la misma para los modelos de inteligencia artificial en este documento, misma tasa de aprendizaje, épocas e iteraciones. La Tabla 4. 4 presenta los resultados obtenidos.

Tabla 4. 4. Criterios estadísticos para la configuración del número de neuronas en la capa oculta.

Arquitectura			Criterios estadísticos								
			Entrenamiento			Prueba			Total		
I	H	O	RMSE	MAPE	R	RMSE	MAPE	R	RMSE	MAPE	R
7	1	1	13.0567	17.3386	0.5372	13.0767	16.764	0.1327	13.0606	17.2256	0.5309
7	2	1	12.457	16.5461	0.5991	11.7583	15.2996	0.2493	12.3298	16.314	0.6048
7	3	1	11.8178	15.7405	0.6506	10.8402	14.8776	0.3893	11.642	15.5798	0.6593
7	4	1	12.9894	17.3329	0.5503	12.4619	16.5198	0.2006	12.8927	17.1815	0.5549
7	5	1	12.5005	16.6859	0.5954	12.031	15.5441	0.2331	12.4144	16.4732	0.5979
7	10	1	12.101	16.1916	0.6292	11.4182	15.0824	0.2884	11.9768	15.985	0.6339
7	20	1	12.1488	15.8984	0.6225	11.5022	15.5233	0.275	12.0266	15.8259	0.6285
<b>7</b>	<b>50</b>	<b>1</b>	<b>11.6873</b>	<b>15.5209</b>	<b>0.6608</b>	<b>11.176</b>	<b>15.4501</b>	<b>0.3446</b>	<b>11.5937</b>	<b>15.5077</b>	<b>0.6626</b>
7	100	1	12.0412	16.1484	0.6338	11.5489	15.1719	0.2862	11.9511	15.9666	0.6353
7	1000	1	13.0107	17.2679	0.5483	12.4481	16.5133	0.2024	12.9078	17.1273	0.5546
Mínimo			11.6873	15.5209	0.6608	10.8402	14.8776	0.1327	11.5937	15.5077	0.5309
Máximo			13.0567	17.3386	0.6608	13.0767	16.764	0.3893	13.0606	17.2256	0.6626
Diferencia			10.5%	10.5%	18.7%	17.1%	11.3%	65.9%	11.2%	10.0%	19.9%

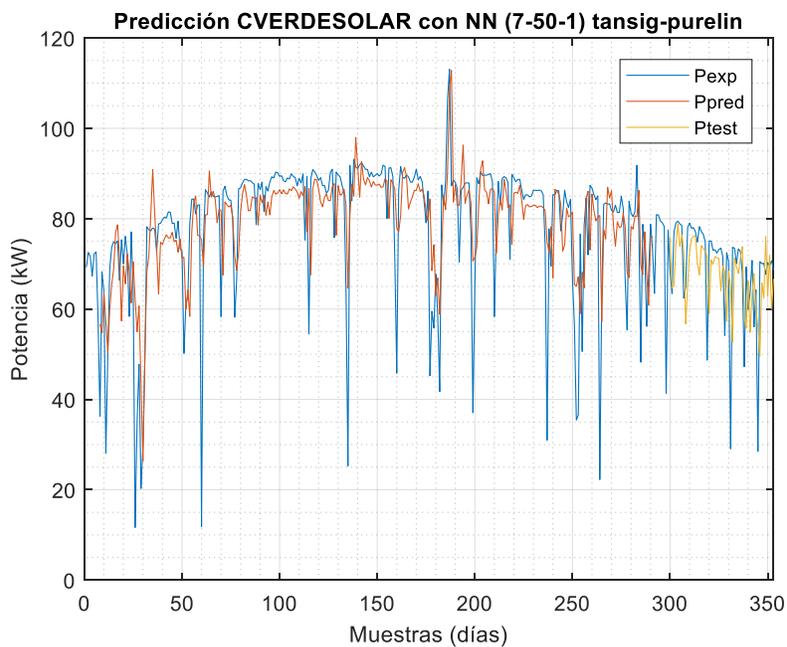


Fig. 4. 10. Comportamiento de la predicción con la configuración de las neuronas en la capa oculta.

Como se observa en la tabla anterior, con 50 neurona alcanzamos los mejores resultados. El incremento de más neuronas no define un mejor desempeño. La gráfica de la Fig. 4. 11 ilustra el comportamiento que sigue la red neuronal clásica entrenada con gradiente descendente.

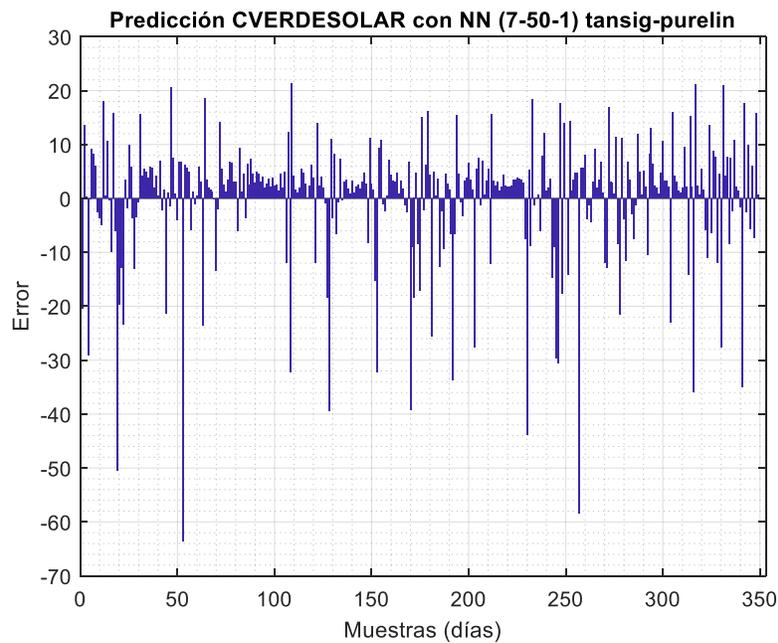


Fig. 4. 11. Error entre la base de datos y la predicción con NN.

## 4.5 Predicción con Redes profundas

Considerando que una red profunda se define por el incremento de capas ocultas y el algoritmo de aprendizaje, se parte de la arquitectura 7-50-1 con tansig de función de transferencia para incrementar las capas y observar si el desempeño es superior a una red clásica. En la Tabla 4. 5 se muestran los resultados de incrementar las capas.

Tabla 4. 5. Criterios Estadísticos de la configuración de profundidad.

Arquitectura			Criterios estadísticos								
			Entrenamiento			Prueba			Total		
I	H	O	RMSE	MAPE	R	RMSE	MAPE	R	RMSE	MAPE	R
7	50(1)	1	11.3978	15.1676	0.6809	11.5273	16.2949	0.3355	11.422	15.3776	0.6763
7	50(2)	1	7.8891	8.7237	0.865	19.3402	24.4758	0.1904	10.9686	11.6574	0.7415
7	50(3)	1	2.9172	2.0862	0.9826	28.0609	33.4927	0.2731	12.3927	7.9356	0.7635
7	50(4)	1	12.5437	16.7145	0.5921	11.9698	15.5252	0.1885	12.4368	16.4888	0.5993
7	50(5)	1	1.562	1.2741	0.995	19.1732	23.5579	0.2398	8.3935	5.4244	0.8628
7	50(10)	1	<b>0.2213</b>	<b>0.2281</b>	<b>0.9999</b>	<b>14.6403</b>	<b>18.8375</b>	<b>0.3614</b>	<b>6.3214</b>	<b>3.6941</b>	<b>0.9181</b>
7	50(20)	1	4.7428	5.133	0.9528	16.709	22.8125	0.2561	8.3847	8.4258	0.8495
Mínimo			0.2213	0.2281	0.5921	11.5273	15.5252	0.1885	6.3214	3.6941	0.5993
Máximo			12.5437	16.7145	0.9999	28.0609	33.4927	0.3614	12.4368	16.4888	0.9181
Diferencia			98.2%	98.6%	40.8%	58.9%	53.6%	47.8%	49.2%	77.6%	34.7%

La diferencia de los errores al incrementar las capas es muy significativa. Mientras que el incremento de neuronas en capas ocultas no mejoran los resultados como se observó en la tabla de redes clásicas, en los entrenamientos con modelos profundos los porcentajes se disparan.

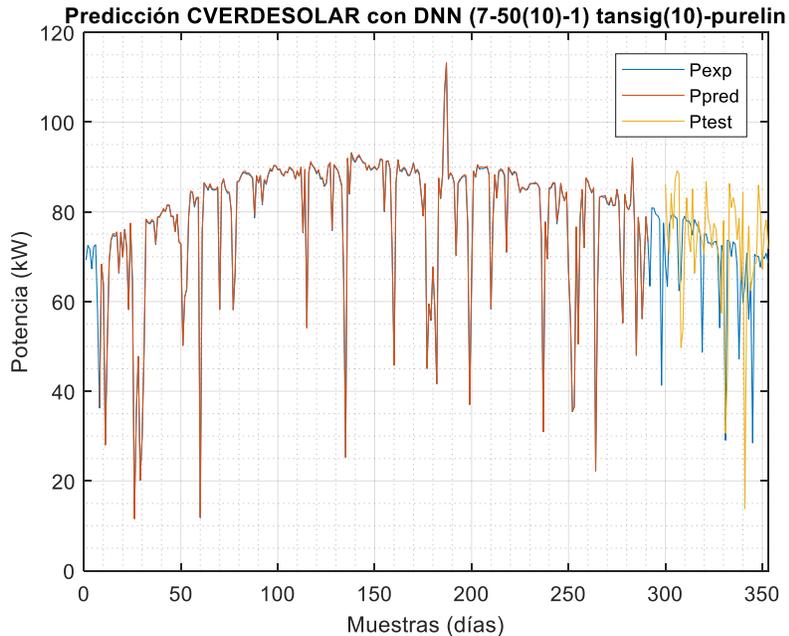


Fig. 4. 12. Comportamiento de la predicción con modelos profundos.

La red profunda con 10 capas ocultas tiene el mejor desempeño con esta base de datos. En la Fig 4. 12 y 4. 13 se observa los comportamientos de las predicciones con respecto a la base de datos original y la diferencia de error, respectivamente.

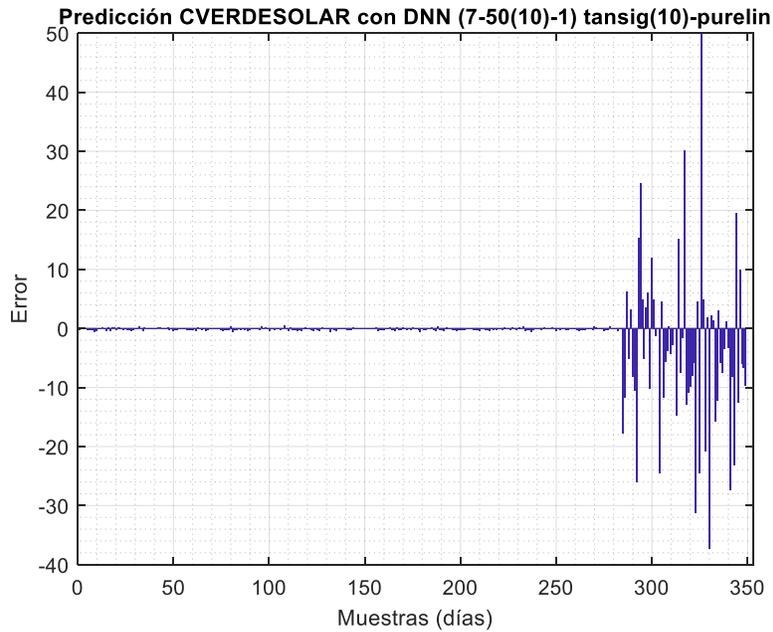


Fig. 4. 13. Error entre la base de datos y el modelo profundo.

## 4.6 Comparación de resultados

Se compararon los resultados para obtener el modelo que mejor ajusta a la base de datos. La Tabla 4. 6 recopila los mejores desempeños para cada modelo. Se observa que el DNN tiene mejores resultados en la etapa de entrenamiento, esto se logró usando el método de

deserción y el algoritmo de ADAM. La red neuronal profunda obtuvo un 0.9999 de correlación a diferencia del NN con 0.6608 y el método estadístico con 0.5612. Cuando se hace la regresión de la base de datos completa, en la columna de Total podemos observar que la red neuronal tiene el mejor desempeño con 0.9181 de correlación, MAPE = 3.6941 y RMSE = 6.3214

Tabla 4. 6. Comparación de los criterios estadísticos de los modelos ajustados.

Modelo	Criterios estadísticos								
	Entrenamiento			Prueba			Total		
	RMSE	MAPE	R	RMSE	MAPE	R	RMSE	MAPE	R
ARMA	12.8351	16.7579	0.5612	8.8117	11.8472	0.6552	12.1376	15.7757	0.6132
NN	11.6873	15.5209	0.6608	11.176	15.4501	0.3446	11.5937	15.5077	0.6626
DNN	0.2213	0.2281	0.9999	14.6403	18.8375	0.3614	6.3214	3.6941	0.9181

---

## Capítulo 5

### Conclusiones

---

Se transformó la base de datos para obtener la potencia por días y adecuar la serie de tiempo para la predicción al día siguiente. Esto significa que en la ventana de predicción las muestras se encontraban por día.

Teniendo una base de datos adecuada para trabajar, se aplicó el método de Cao para determinar la ventana de embebimiento, resultando este método bastante acertado para determinar la cantidad de neuronas de entrada para la predicción con redes neuronales.

En orden para determinar la mejor red neuronal, se variaron las funciones de activación, capas y número de neuronas, repitiendo el procedimiento de entrenamiento hasta conseguir resultados que no presentaron variaciones en sus errores cuadráticos en cifras decimales.

Los modelos profundos presentaron un desempeño superior según los criterios estadísticos involucrados en este documento para nuestro problema en particular. Estos resultados son válidos para la ubicación donde se registró la información de la base de datos de potencia eléctrica de la planta proporcionada. A pesar de los buenos resultados, se observó que existe una especialización en la red cuando se continúan incrementando las capas

cultas. Esto puede ser debido a que los modelos profundos requieren una base de datos más grande para realizar sus ajustes.

El modelo ARMA tiene el mejor desempeño cuando se aplicó la predicción a la base de datos de prueba, lo que significa que para hacer predicciones teniendo una base de datos con muestras pasadas de un año, estos modelos matemáticos pueden ser suficientes para predecir a corto plazo.

Las redes neuronales clásicas presentaron un pobre ajuste pero una especialización menor, lo que puede ser tomado como ventaja para las predicciones con pocas muestras.

Para trabajos futuros se puede tomar una base de datos más amplia con la finalidad de mejorar las predicciones con modelos profundos a largo plazo.

# Referencias

---

## Bibliografía

- Attaviriyapap, P., Tokuhara, K., Marmioli, M., Tsukamoto, Y., & Kojima, Y. (2011). Estimation of photovoltaic power generation output based on solar irradiation and frequency classification. *Innovative Smart Grid Technologies Asia (ISGT) 2011 IEEE/PES*, 1-5.
- Bouzerdoun, M., Mellit, A., & Massi Pavan, A. (2013). A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy*, 98, 226-235.
- Brockwell, P., & Davis, R. (2002). *Introduction to Time Series and Forecasting*. New York, USA: Springer-Verlag.
- Cao, L. (1997). Practical method for determining the minimum embedding dimension of a scalar time series. *Physical D*, Vol. 110, 43-50.
- CFE. (MAYO de 2010). MEDIDORES MULTIFUNCIÓN PARA SISTEMAS ELÉCTRICOS CFE G0000-48. México.
- Chu, Y., Urquhart, B., Gohari, S., Hugo, P., Kleissl, J., & Coimbra, C. (2015). Short-term reforecasting of power output from a 48 MWe solar PV plant. *Solar Energy*, 112, 68-77.

- Cococcioni, M., D'Andrea, E., & Lazzerini, B. (2011). 24-hour-ahead forecasting of energy production in solar PV systems. *2011 11th International Conference on Intelligent Systems Design and Applications (ISDA)* (págs. 1-6). Cordoba, Spain: IEEE.
- COMISIÓN REGULADORA DE ENERGÍA. (8 de ABRIL de 2016). RESOLUCIÓN Núm. RES/151/2016. México.
- Diario Oficial de la Federación. (02 de Julio de 2015). Criterios mediante los que se establecen las características específicas de la infraestructura requerida para la Interconexión de Centrales Eléctricas y Conexión de Centros de Carga. México.
- Duffie, J., & Beckman, W. (2005). *Solar Engineering of Thermal Process*. New Jersey: Jhon Wiley & Sons.
- Dunlop, J. (2010). *Photovoltaic Systems*. Illinois, EUA: American Technical Publishers.
- Espadas, G. (2002). *Reconocimiento de patrones usando redes neuronales artificiales*. Mérida, Yucatán, México.
- Fausett, L. (1994). *Fundamentals of neural networks: architectures, algorithms, and applications*. New Jersey, EUA: Uppler Saddle: Prentice-Hall.
- Fernandez-Jimenez, L., Muñoz-Jimenez, A., Falces, A., Mendoza-Villena, M., García-Garrido, E., Lara-Santanilla, P., . . . Zorzano-Santamaría, P. (2012). Short-term power forecasting system for photovoltaic plants. *Renewable Energy*, *44*, 311-317.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. USA: Elsevier.

- Henan, L. (2013). Neural network based learning method for estimating power output from forecasted irradiance for solar photovoltaic system. *Tesis de Maestría, Ingeniería Eléctrica*. Texas, EUA.
- Honglu Zhu, Xu Li, Qiao Sun, Ling Nie, Jianxi Yao, & Gang Zhao. (2015). A Power Prediction Method for Photovoltaic Power Plant Based on Wavelet Decomposition and Artificial Neural Networks. *Energies*, 1-15.
- Huaizhi Wang, Haiyan Yi, Jianchun Peng, Guibin Wang, Yitao Liu, Hui Jiang, & Wenxin Liu. (2017). Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. *Energy Conversion and Management*, 153, 409-422.
- Hugo, P., & Coimbra, C. (2012). Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, 86, 2017-2028.
- Judkins, R., Fulkerson, W., & Sanghvi, M. (1993). The dilemma of fossil fuel use and global climate change. *Energy Fuels* 7, Vol. 1, 14-22.
- Karamirad, M., Omid, M., Alimardani, R., Mousazadeth, H., & Heidari, S. (2013). ANN based simulation and experimental verification of analytical four- and five-parameters models of PV modules. *Simulation Modelling Practice and Theory*, 34, 86-98.
- Kingma, D., & Ba, J. (2015). ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION. *ICLR*.
- LAERFTE. (28 de Noviembre de 2008). LEY PARA EL APROVECHAMIENTO DE ENERGÍAS RENOVABLES Y EL FINANCIAMIENTO DE LA TRANSICIÓN ENERGÉTICA. México: Diario Oficial de la Federación.

- Leva, S., Dolara, A., Grimaccia, F., Mussetta, M., & Ogliari, E. (2017). Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. *Mathematics and Computers in Simulation*, 131, 88-100.
- Lynn, P. (2010). *Electricity from Sunlight: An Introduction to Photovoltaics*. London, UK: John Wiley & Sons.
- Mekki, H., Salhi, H., & Khaled, B. (2007). Modeling and Simulation of Photovoltaic Panel Based on Artificial Neural Networks and VHDL-language. *14th IEEE International Conference on: Electronics, Circuits and Systems*.
- Oudjana, S., Hellal, A., & Mahammed, I. (2013). Power Forecasting of Photovoltaic Generation. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*, Vol. 7.
- Picault, D., Raison, B., Bacha, S., de la Casa, J., & Aguilera, J. (2010). Forecasting photovoltaic array power production subject to mismatch losses. *Solar Energy*, 84, 1301-1309.
- Polson, N., & Sokolov, V. (2017). Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, Vol. 79, 1-17.
- Quaschnig, V. (2005). *Understanding Renewable Energy*. London, UK: Earthscan.
- Sahnoun, M., Romero, H., Carmona, J., & Gomand, J. (2013). Maximum power point tracking using P&O control optimized by a neural network approach: a good compromise between accuracy and complexity. *Energy Procedia*, 42.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, Vol. 61, 85-117.

- Secretaría de Energía. (29 de Febrero de 2012). PROSPECTIVA DE ENERGÍAS RENOVABLES 2012-2026. México.
- Shiva Kumar, B., & Sudhakar, K. (2015). Performance evaluation of 10 MW grid connected solar photovoltaic power plant in India. *Energy Reports*, 1, 184-192.
- Small, M. (2005). *Applied Nonlinear Time Series Analysis: Applications In Physics, Physiology And Finance*. Singapore: World Scientific.
- Takens, F. (1981). Lecture Notes in Mathematics. Vol. 366.
- Tziu Tzib, J., Alejos Moo, E., Bassam, A., Flota-Bañuelos, M., Escalante Soberanis, M., Ricalde, M., & López-Sánchez, M. (2016). Photovoltaic Module Temperature Estimation: A Comparison Between Artificial Neural Network and Adaptive Neuro Fuzzy Inference Systems Models. *Intelligent Computing Systems*, 597.
- Ud Din, G., & Marnerides, A. (2017). Short term power load forecasting using Deep Neural Networks. *2017 International Conference on Computing, Networking and Communications (ICNC): Green Computing, Networking, and* (págs. 1-5). Santa Clara, CA, USA: IEEE.
- WEC. (2013). *Survey of Energy Resources 2013*. Londres, Inglaterra: World Energy Council.
- Xueheng Qiu, Le Zhang, Ye Ren, P.N. Suganthan, & Gehan Amaratunga. (2014). Ensemble deep learning for regression and time series forecasting. *2014 IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL)* (págs. 1-6). Orlando, FL, USA: IEEE.