



# UNIVERSIDAD TÉCNICA DE COTOPAXI

## DIRECCIÓN DE POSGRADO

### MAESTRÍA EN ELECTRICIDAD

### MODALIDAD: PROYECTO DE INVESTIGACIÓN

**Título:**

---

ANÁLISIS DEL POTENCIAL DE IRRADIACIÓN SOLAR EN LA  
HACIENDA "LA CAMPIÑA" DE LA PARROQUIA DE MULALÓ.  
PROPUESTA PARA APLICACIÓN DE UN ALGORITMO DE  
OPTIMIZACIÓN BASADO EN MACHINE LEARNING EN LA PREDICCIÓN  
DE GENERACIÓN FOTOVOLTAICA.

---

Trabajo de titulación previo a la obtención del título de Magister en Electricidad  
mención Sistemas Eléctricos de Potencia

**Autor:**

Núñez Verdezoto Marlon Daniel

**Tutor:**

M. Sc. Álvaro Santiago Mullo Quevedo

LATACUNGA –ECUADOR

2022

## **AVAL DEL TUTOR**

En mi calidad de Tutor del Trabajo de Titulación. Análisis del potencial de irradiación solar en la Hacienda "La Campiña" de la parroquia de Mulaló. Propuesta para aplicación de un algoritmo de optimización basado en Machine Learning en la predicción de generación fotovoltaica. Presentado por Núñez Verdezoto Marlon Daniel, para optar por el título Magíster en Electricidad Mención Sistemas Eléctricos de Potencia.

## **CERTIFICO**

Que dicho trabajo de investigación ha sido revisado en todas sus partes y se considera que reúne los requisitos y méritos suficientes para ser sometido a la presentación para la valoración por parte del Tribunal de Lectores que se designe y su exposición y defensa pública.

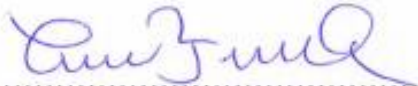
Latacunga, Agosto, 16, 2022

  
.....  
M. Sc. Alvaro Santiago Mullo Quevedo  
C.I.: 050276854-2

## AVAL DEL TRIBUNAL

El trabajo de Titulación: Análisis del potencial de irradiación solar en la Hacienda "La Campiña" de la parroquia de Mulaló. Propuesta para aplicación de un algoritmo de optimización basado en Machine Learning en la predicción de generación fotovoltaica, ha sido revisado, aprobado y autorizado su impresión y empastado, previo a la obtención del título de Magíster en Electricidad Mención Sistemas Eléctricos de Potencia; el presente trabajo reúne los requisitos de fondo y forma para que el estudiante pueda presentarse a la exposición y defensa.

Latacunga, noviembre, 15, 2022



.....  
M. Sc. Carlos Iván Quinatoa Caiza  
C.I. 0503287864  
Presidente del tribunal



.....  
M. Sc. Franklin Vásquez Teneda  
C.I. 1710434497  
Lector 2



.....  
M. Sc. Carlos Francisco Pacheco Mena  
C.I. 0503072902  
Lector 3

## **DEDICATORIA**

El siguiente trabajo de investigación va dedicado a todas las personas que han formado parte de mi vida apoyándome en cada decisión, ayudándome a crecer y creer en mí.

Dedico principalmente a mi madre con su esfuerzo, su ejemplo se convirtió en mi guía y mi ejemplo de seguir siempre para adelante.

Marlon Daniel Núñez Verdezoto

## AGRADECIMIENTO

En primer lugar, agradezco a Dios por la vida brindada y darme sabiduría para culminar esta etapa como profesional y guiarme siempre por el camino del bien.

Agradecido con mi madre BERTHA NÚÑEZ, mi tía ALICIA NÚÑEZ quienes, con sus consejos, enseñanza, valores inculcados en mi crecimiento, me han enseñado a no rendirme a seguir luchar por los objetivos planteados, mediante su ejemplo me he convertido en una persona de bien.

Agradecerle a la Universidad Técnica de Cotopaxi, una prestigiosa institución educativa que me abrió sus puertas, siendo la base de mi formación y testigo de mi preparación profesional desde pregrado hasta posgrado.

Eternamente agradecido con mis docentes en especial con mi tutor y amigo ÁLVARO MULLO una excelente persona y profesional que ha confiado en mí y me ha brindado su apoyo tanto profesional como personal desde mi formación como ingeniero, y en la obtención de mi maestría.

Para finalizar agradecer a mis compañeros que, con sus ocurrencias, enseñanzas en este tiempo se fueron convirtiendo en unos grandes amigos.

Marlon Daniel Núñez Verdezoto

## **RESPONSABILIDAD DE AUTORÍA**

Quien suscribe, declara que asume la autoría de los contenidos y los resultados obtenidos en el presente trabajo de titulación.

Latacunga, agosto, 16, 2022

  
.....  
Marlon Daniel Núñez Verdezoto  
C.I. 020235043-5

## RENUNCIA DE DERECHOS

Quien suscribe, cede los derechos de autoría intelectual total y/o parcial del presente trabajo de titulación a la Universidad Técnica de Cotopaxi.

Latacunga, agosto, 16, 2022

  
.....  
Marlon Daniel Núñez Verdezoto  
C.I. 020235043-5

## **AVAL DEL PRESIDENTE DEL TRIBUNAL**

Quien suscribe, declara que el presente Trabajo de Titulación: Análisis del potencial de irradiación solar en la hacienda "La Campiña" de la parroquia de Mulaló. Propuesta para aplicación de un algoritmo de optimización basado en Machine Learning en la predicción de generación fotovoltaica, contiene las correcciones a las observaciones realizadas por los lectores en sesión científica del tribunal.

Latacunga, noviembre, 15, 2022



M. Sc. Carlos Iván Quinatoa Caiza  
C.I. 0503287864



**UNIVERSIDAD TÉCNICA DE COTOPAXI  
DIRECCIÓN DE POSGRADO**

**MAESTRÍA EN ELECTRICIDAD  
MENCIÓN SISTEMAS ELÉCTRICOS DE POTENCIA**

**Título:** Análisis del potencial de irradiación solar en la Hacienda "La Campiña" de la parroquia de Mulaló. Propuesta para aplicación de un algoritmo de optimización basado en Machine Learning en la predicción de generación fotovoltaica.

**Autor:**

Núñez Verdezoto Marlon Daniel

**Tutor:**

Mullo Quevedo Álvaro Santiago, M. Sc

**RESUMEN**

La investigación presenta un modelo de predicción de generación fotovoltaica utilizando Machine Learning y una de sus principales máquinas de aprendizaje "Random Forest", esta investigación tuvo lugar en La Hacienda "La Campiña" ubicada en la parroquia Mulaló que pertenece al cantón Latacunga, donde se realizó la toma de datos reales de Irradiación solar y Temperatura en el Lapso del año 2020 con un instrumento de medición llamado piranómetro Solar Power Meter SM206. El modelo de predicción fue creado en Python y las variables de mayor incidencia. Este modelo realiza una predicción de la variable Potencia, fue entrenado con las variables de Irradiación solar y Temperatura el proceso consiste en ingresar las variables primero de un día cualquiera para predecir una variable seleccionada de otro día, aquí tomamos datos del día 01 de enero para su entrenamiento y predecimos el día 20 del mismo mes con la mayor cantidad de datos de variables para tener un error que se encuentra entre los márgenes permisibles y garantice la eficiencia del modelo de la misma manera se realizó la predicción de un mes completo donde se entrenó con datos del mes marzo para llegar a predecir mayo. La eficiencia en la predicción diaria es del 91,41% y la mensual del 94,47%, la predicción mensual llega a ser más efectiva debido a que contiene mayor número de datos de cada variable por ende el desempeño en el entrenamiento es más rentable, cabe recalcar que este modelo utiliza 1 000 Árboles de decisión.

**PALABRAS CLAVE:** Sistema Fotovoltaico, Machine Learning, Algoritmo, Árbol de decisión, Predicción, Variable estocástica.

**UNIVERSIDAD TECNICA DE COTOPAXI  
DIRECCION DE POSGRADO**

**MAESTRÍA EN ELECTRICIDAD  
MENCION SISTEMAS ELÉCTRICOS DE POTENCIA**

**Topic:** Solar irradiation potential analysis in the "La Campiña" farming from Mulaló parish. Proposal for the optimization algorithm application based on Machine Learning in the photovoltaic generation prediction.

**Author:**

Núñez Verdezoto Marlon Daniel

**Tutor:**

M.Sc. Mullo Quevedo Álvaro Santiago

**ABSTRACT**

The research presents a photovoltaic generation prediction model using Machine Learning and one its main "Random Forest" learning machines, this research took place in the "La Campiña" farming, located in the Mulaló parish, what belonging to the Latacunga canton, where carried out the solar irradiation real data collection and temperature in the 2020 year period with a measuring instrument called Solar Power Meter SM206 piranómetro. The prediction model was created in Python and the variables with the highest incidence. This model makes a Power variable prediction, it was trained with the Solar Irradiation variables and Temperature, the process consists of entering the variables, first of any given day to predict a another day chosen variable, here, it takes data from January 1 to its training and it predicts the same month 20th with the variable data largest amount to have an error, which is within the permissible margins and ensure the model efficiency, in the same way, it was made the a complete month prediction, where it trained with data from the March month to predict May. The efficiency in the daily prediction is 91.41% and the monthly 94.47%, the monthly prediction becomes more effective, because it contains a data greater number each variable, therefore, the training performance is more profitable, it should be noted, which this model uses 1000 decision trees.

**KEYWORDS:** Photovoltaic system, machine learning, algorithm, random forest, prediction, stochastic variable.

Yo, Marco Paul Beltrán Semblantes con cédula de identidad número:0502666514 Magister en Lingüística Aplicada a la Enseñanza del idioma Inglés como Lengua Extranjera con número de registro de la SENESCYT: 1020-2021-2354162; **CERTIFICO** haber revisado y aprobado la traducción al idioma Inglés del resumen del trabajo de investigación con el título: "Análisis del potencial de irradiación solar en la Hacienda "La Campiña" de la parroquia de Mulaló. Propuesta para aplicación de un algoritmo de optimización basado en Machine Learning en la predicción de generación fotovoltaica" de: Núñez Verdezoto Marlon Daniel, aspirante a Magister en Electricidad, Mención Sistemas Eléctricos de Potencia.



Mg. Marco Paul Beltrán Semblantes  
0502666514



CENTRO  
DE IDIOMAS

Latacunga, agosto, 2022

## ÍNDICE DE CONTENIDO

<b>AVAL DEL TUTOR</b> .....	<b>ii</b>
<b>AVAL DEL TRIBUNAL</b> .....	<b>iii</b>
<b>DEDICATORIA</b> .....	<b>iv</b>
<b>AGRADECIMIENTO</b> .....	<b>v</b>
<b>RESPONSABILIDAD DE AUTORÍA</b> .....	<b>vi</b>
<b>RENUNCIA DE DERECHOS</b> .....	<b>vii</b>
<b>AVAL DEL PRESIDENTE DEL TRIBUNAL</b> .....	<b>viii</b>
<b>RESUMEN</b> .....	<b>ix</b>
<b>ÍNDICE DE CONTENIDO</b> .....	<b>xi</b>
<b>ÍNDICE DE ANEXOS</b> .....	<b>xvii</b>
<b>INTRODUCCIÓN</b> .....	<b>1</b>
Antecedentes: .....	1
Planteamiento del problema: .....	3
Formulación del problema:.....	3
Objetivo General: .....	4
Objetivos Específicos:.....	4
Sistemas de tareas en relación a los objetivos específicos: .....	4
Justificación.....	6
<b>CAPÍTULO I</b> .....	<b>7</b>
<b>FUNDAMENTACIÓN TEÓRICA- METODOLÓGICA</b> .....	<b>7</b>
1.1    Antecedentes de la investigación o fundamentación del estado del arte..	7
1.2    Fundamentación Teórica .....	8
1.2.1    Sistemas fotovoltaicos.....	8
1.2.2    Clasificación de los Sistemas fotovoltaicos .....	12
1.2.3    Radiación solar.....	14

1.2.4	Irradiación Solar.....	15
1.2.5	Horas pico solares .....	18
1.2.6	Inteligencia artificial .....	20
1.3	Fundamentación metodológica.....	34
1.3.1	Tipo de investigación .....	34
1.3.2	Método de investigación .....	34
1.3.3	Técnicas e instrumentos .....	35
	Conclusiones del capítulo I .....	36
	<b>CAPÍTULO II.....</b>	<b>37</b>
<b>2</b>	<b>PROPUESTA.....</b>	<b>37</b>
2.1	Título del proyecto .....	37
2.2	Objetivos del proyecto.....	37
2.3	Justificación.....	37
2.4	Descripción de la metodología .....	38
2.5	Metodología para cumplir los objetivos planteados.....	38
2.5.1	Histograma y frecuencia de irradiación solar.....	42
2.5.2	Diagrama de Pareto de la irradiación solar .....	43
2.5.3	Análisis de la potencia requerida en el sector de la investigación ..	44
2.5.4	Factor de planta de la hacienda “La Campiña” .....	47
2.5.5	Matriz de confusión para selección del algoritmo .....	48
2.5.6	Flujograma y pseudocódigo empleado en el análisis del proceso... 49	
2.5.7	Análisis de la división de DataFrame de entrenamiento, validación y prueba. 52	
2.5.8	Modelo de la Red Neuronal (Random Forest) .....	53
2.5.9	Entrenamiento del modelo de predicción.....	56
2.5.10	Cálculo del Error Absoluto .....	59

Conclusiones del capítulo II .....	61
<b>CAPÍTULO III.....</b>	<b>62</b>
3.1 Análisis de resultados .....	62
3.1.1 Análisis de las horas con mayor incidencia en el día.....	63
3.1.2 Toma de valores de potencia en la hacienda.....	64
3.1.3 Inicio y adecuación del programa .....	67
3.1.4 Resultados de la ejecución del modelo de predicción diario .....	69
3.1.5 Resultados de la validación del modelo de predicción mensual .....	71
3.1.6 Validación cruzada del modelo de predicción (Cross-Validation) .	74
Conclusiones del capítulo III.....	77
<b>CONCLUSIONES.....</b>	<b>78</b>
<b>RECOMENDACIONES.....</b>	<b>80</b>
<b>REFERENCIAS BIBLIOGRAFÍA .....</b>	<b>81</b>
<b>ANEXOS .....</b>	<b>83</b>

## ÍNDICE DE FIGURAS

FIGURA 1.1. ESPECTRO LUMINOSO [11]. -----	9
FIGURA 1.2. CELDAS DE SILICIO MONOCRISTALINO, POLICRISTALINO Y AMORFO [6]. -----	10
FIGURA 1.3. BOSQUEJO DE UN SISTEMA SOLAR FOTOVOLTAICO [2]. -----	12
FIGURA 1.4. SISTEMA SOLAR FOTOVOLTAICO AUTÓNOMO [12]. -----	13
FIGURA 1.5. ESQUEMA DE UN SFCR [13]. -----	14
FIGURA 1.6. RADIACIÓN SOLAR [6]. -----	15
FIGURA 1.7. COMPROBACIÓN DE DATOS DE IRRADIACIÓN RECOLECTADOS. -----	16
FIGURA 1.8. ATLAS SOLAR ELABORADO POR CONELEC [16].-----	17
FIGURA 1.9. IRRADIACIÓN HORIZONTAL GLOBAL SOLARGIS. -----	17
FIGURA 1.10. IRRADIACIÓN NORMAL DIRECTA SOLARGIS.-----	18
FIGURA 1.11. PORCENTAJE DE CAÍDA DE IRRADIACIÓN SOLAR POR EMISIÓN DE CENIZA. -----	19
FIGURA 1.12. DIFERENCIA ENTRE LA INTELIGENCIA ARTIFICIAL Y LOS DEMÁS PARÁMETROS [19]. -----	21
FIGURA 1.13. APRENDIZAJE DE MÁQUINAS. -----	22
FIGURA 1.14. RED NEURONAL DE TRES CAPAS [19].-----	23
FIGURA 1.15. ALGORITMO DE AGRUPACIÓN CON DATOS BIDIMENSIONALES [19]. --	24
FIGURA 1.16. DIAGRAMA DE FLUJO CON LOS PASOS DE K-MEANS [19]. -----	27
FIGURA 1.17. ESTRUCTURA DE UN ÁRBOL DE DECISIÓN [21]. -----	30
FIGURA 1.18. TEORÍA DE K-VECINOS MÁS CERCANOS (LIGDI GONZALES). -----	31
FIGURA 1.19. RED NEURONAL ARTIFICIAL [20]. -----	32
FIGURA 1.20. RED NEURONAL ANW [20].-----	33
FIGURA 1.21. CÁLCULO DEL ERROR EN PREDICCIÓN [19]. -----	33
FIGURA 2.1. LUGAR DE LA TOMA DE DATOS. GOOGLE EARTH.....	38
FIGURA 2.2. CURVA DE IRRADIACIÓN DE ENERO.....	39
FIGURA 2.3. CURVA DE IRRADIACIÓN SOLAR DE UN DÍA DEL MES DE ENERO .....	40
FIGURA 2.4. CURVA DE VARIACIÓN DE IRRADIACIÓN PROMEDIO DEL DÍA. ....	40
FIGURA 2.5. PROMEDIO DE IRRADIACIÓN POR MES.....	41

FIGURA 2.6. COMPORTAMIENTO DE LA IRRADIACIÓN SOLAR PROMEDIO DEL AÑO..	42
FIGURA 2.7. HISTOGRAMA. ....	43
FIGURA 2.8. DIAGRAMA DE PARETO. ....	44
FIGURA 2.9. CURVA DE CARGA. ....	45
FIGURA 2.10. MATRIZ DE CONFUSIÓN CON ÁRBOLES DE DECISIÓN.....	49
FIGURA 2.11. DIAGRAMA DE FLUJO DEL PROCESO.....	50
FIGURA 2.12. DATOS DE ENTRENAMIENTO DEL DÍA 1 DE ENERO. ....	52
FIGURA 2.13. TEMPERATURA DE ENTRENAMIENTO DEL DÍA 1 DE ENERO. ....	53
FIGURA 2.14. PROCESO DE ENTRENAMIENTO. ....	53
FIGURA 2.15. VARIABLES CON MAYOR INCIDENCIA. ....	54
FIGURA 2.16. PROCESO DE CODIFICACIÓN DE DATOS O VARIABLES.....	55
FIGURA 2.17. PROCESO DE CODIFICACIÓN DE DATOS O VARIABLES.....	57
FIGURA 2.18. CURVAS DEL MES DE MARZO.....	58
FIGURA 2.19. PROCESO DE CODIFICACIÓN DE DATOS O VARIABLES MENSUALES ....	59
FIGURA 3.1. CURVA DE COMPARACIÓN DE DATOS.....	63
FIGURA 3.2. DIAGRAMA (FRECUENCIA VS W/M <sup>2</sup> ) DE LA MAYOR INCIDENCIA DE IRRADIACIÓN SOLAR .....	64
FIGURA 3.3. DIAGRAMA (FRECUENCIA W/M <sup>2</sup> AL DÍA) .....	65
FIGURA 3.4. DATOS OBTENIDOS EN LA HACIENDA “LA CAMPIÑA”.....	66
FIGURA 3.5. LIBRERÍAS UTILIZADAS. ....	67
FIGURA 3.6. PSEUDOCÓDIGO DE VARIABLES. ....	69
FIGURA 3.7. PSEUDOCÓDIGO DE GRÁFICO DE POTENCIA ORIGINAL VS POTENCIA PREDICHA.....	69
FIGURA 3.8. GRÁFICA DE RESULTADO.....	70
FIGURA 3.9. GRÁFICA DE RESULTADO.....	70
FIGURA 3.10. ERROR ABSOLUTO. ....	71
FIGURA 3.11. VARIACIÓN DE COMPONENTES PRINCIPALES. ....	72
FIGURA 3.12. PSEUDOCÓDIGO DE LAS CURVAS. ....	73
FIGURA 3.13. POTENCIA ORIGINAL VS POTENCIA DE PREDICCIÓN.....	73
FIGURA 3.14. ERROR DE PREDICCIÓN MENSUAL.....	74
FIGURA 3.15. CURVAS DE DESEMPEÑO.....	76

## ÍNDICE DE TABLAS

TABLA 1.1. RENDIMIENTO DE PANELES POR SU MATERIAL DE FABRICACIÓN. -----	10
TABLA 1.2. IRRADIACIÓN SOLAR COMPROBADA CON PANELES DE DIFERENTE POTENCIA.-----	16
TABLA 2.1. DATOS PROMEDIO ANUAL DE IRRADIACIÓN .....	41
TABLA 2.2. POTENCIA DE CARGA ACTUAL DEL SITIO.....	45
TABLA 2.3. MEDIDAS DE RENDIMIENTO .....	48
TABLA 3.1. COMPARACIÓN CON DATOS MEDIDOS .....	62
TABLA 3.2. DATASHEET DEL PANEL FOTOVOLTAICO. ....	64
TABLA 3.3. POTENCIA ACUMULADA APROXIMADA EN UN MES.....	65
TABLA 3.4. CÁLCULO DE LA ENERGÍA ASUMIENDO UN PANEL FOTOVOLTAÍCO.....	66
TABLA 3.5. DATOS QUE SE CARGARON AL MODELO. ....	68
TABLA 3.6. DATOS ERRÓNEOS DE LA EJECUCIÓN DEL PROGRAMA.....	68
TABLA 3.7. TABLA DE CÁLCULO DEL ERROR. ....	71
TABLA 3.8. TABLA DE EVALUACIÓN DE VALIDACIÓN CRUZADA.....	75



## ÍNDICE DE ANEXOS

ANEXO 1: CURVA DE IRRADIACIÓN SOLAR DE FEBRERO.....	83
ANEXO 2: IRRADIACIÓN SOLAR DIARIA PROMEDIO DE FEBRERO.....	83
ANEXO 3: CURVA DE IRRADIACIÓN SOLAR DE MARZO. ....	84
ANEXO 4: IRRADIACIÓN SOLAR DIARIA PROMEDIO DE MARZO. ....	84
ANEXO 5: CURVA DE IRRADIACIÓN SOLAR DE ABRIL. ....	85
ANEXO 6: IRRADIACIÓN SOLAR DIARIA PROMEDIO DE ABRIL. ....	85
ANEXO 7: CURVA DE IRRADIACIÓN SOLAR DE MAYO. ....	86
ANEXO 8: IRRADIACIÓN SOLAR DIARIA PROMEDIO DE MAYO. ....	86
ANEXO 9: CURVA DE IRRADIACIÓN SOLAR DE JUNIO.....	87
ANEXO 10: IRRADIACIÓN SOLAR DIARIA PROMEDIO DE JUNIO.....	87
ANEXO 11: CURVA DE IRRADIACIÓN SOLAR DE JULIO.....	88
ANEXO 12: IRRADIACIÓN SOLAR DIARIA PROMEDIO DE JULIO.....	88
ANEXO 13: CURVA DE IRRADIACIÓN SOLAR DE AGOSTO.....	89
ANEXO 14: IRRADIACIÓN SOLAR DIARIA PROMEDIO DE AGOSTO.....	89
ANEXO 15: CURVA DE IRRADIACIÓN SOLAR DE SEPTIEMBRE.....	90
ANEXO 16: IRRADIACIÓN SOLAR DIARIA PROMEDIO DE SEPTIEMBRE.....	90
ANEXO 17: CURVA DE IRRADIACIÓN SOLAR DE OCTUBRE. ....	91
ANEXO 18: IRRADIACIÓN SOLAR DIARIA PROMEDIO DE OCTUBRE. ....	91
ANEXO 19: CURVA DE IRRADIACIÓN SOLAR DE NOVIEMBRE. ....	92
ANEXO 20: IRRADIACIÓN SOLAR DIARIA PROMEDIO DE NOVIEMBRE. ....	92
ANEXO 21: CURVA DE IRRADIACIÓN SOLAR DE DICIEMBRE. ....	93
ANEXO 22: IRRADIACIÓN SOLAR DIARIA PROMEDIO DE DICIEMBRE. ....	93
ANEXO 23: IRRADIACIÓN PROMEDIO DE CADA MES.....	94
ANEXO 24: CURVA DE IRRADIACIÓN PROMEDIO DE CADA MES. ....	94
ANEXO 25: TABLA DE FRECUENCIAS DE IRRADIACIÓN SOLAR.....	95
ANEXO 26: TORTA DE FRECUENCIAS DE IRRADIACIÓN SOLAR.....	95
ANEXO 27: TABLA DE PRUEBA DEL MODELO DE PREDICCIÓN.....	95
ANEXO 28: CURVA DE IRRADIACIÓN DEL DÍA 20. ....	96
ANEXO 29: CURVA DE TEMPERATURA DEL DÍA 20.....	96
ANEXO 30: CÓDIGO PARA GRAFICAR CURVAS.....	97
ANEXO 31: PSEUDOCÓDIGO DE IMPORTACIÓN DE DATOS FOTOVOLTAICOS. ....	97

ANEXO 32: TABLA DE IMPORTACIÓN DE DATOS FOTOVOLTAICOS MENSUALES. .	98
ANEXO 33: TABLA DIAGRAMA DE FRECUENCIAS. ....	98
ANEXO 34: DIAGRAMA (FRECUENCIA VS W/M2) DE LA MAYOR INCIDENCIA DE IRRADIACIÓN SOLAR. ....	99
ANEXO 35: TABLA DIAGRAMA DE FRECUENCIAS DE MENOR INCIDENCIA DE IRRADIACIÓN SOLAR. ....	99
ANEXO 36: PANEL FOTOVOLTAICO ECOGREE ENERGY. ....	100
ANEXO 37: DATASHEET DEL PANEL FOTOVOLTAICO ECOGREE ENERGY. ...	101
ANEXO 38: TABLA DE VARIABLES UTILIZADAS. ....	102
ANEXO 39: CURVA DE IRRADIACIÓN SOLAR MES DE MAYO. ....	102
ANEXO 40: CURVA DE TEMPERATURA MES DE MAYO. ....	103
ANEXO 41: PRUEBAS DE PREDICCIÓN. ....	103
ANEXO 42: CURVA DE TEMPERATURA MES DE MAYO. ....	104
ANEXO 43: CURVA DE TEMPERATURA MES DE MAYO. ....	104
ANEXO 44: CÁLCULO DE LA POTENCIA. ....	105
ANEXO 45: RESULTADOS ARROJADOS POR EL PROGRAMA. ....	108
ANEXO 46: MATRIZ DE CONFUSIÓN CON K-MEANS. ....	108
ANEXO 47: MATRIZ DE CONFUSIÓN CON REDES NEURONALES. ....	109

# INTRODUCCIÓN

## **Antecedentes:**

Como se conoce la electricidad es indispensable en la humanidad, países han logrado desarrollos increíbles gracias a la energía eléctrica. La generación convencional de electricidad se ha venido dando con el uso de energías fósiles, pero es materia que produce contaminación y las reservas se comienzan a agotar, es por ello que se ha dado luz verde a las energías alternativas que se encuentran de una manera infinita en el planeta, entre estas energías existe la fotovoltaica que es emitida por el sol en forma de irradiación, con metodologías y estudios actuales se está aprovechando aunque van apareciendo problemáticas para ser utilizada de una manera adecuada [1] [2].

Mediante este previo análisis se llega a la problemática existente en haciendas ya que se construyen a distancia largas de la población con un fin turístico, esto conlleva a la aparición de inconvenientes para abastecerse de servicios básicos entre ellos la electricidad que cumple un rol muy importante por su eficiencia y disponibilidad, ya que brinda iluminación y energía para la producción de alimentos. Para cubrir esta problemática se ha comenzado a invertir en energías alternativas por su facilidad de implementación y el costo es más bajo que llevar energía eléctrica desde una fuente alimentadora [3].

La irradiación solar llega de una manera continua al planeta pero no es constante es decir va variando de acuerdo a la hora del día o la estación que nos encontremos, al realizar modelos de sistemas fotovoltaicos no se toma muy en cuenta los parámetros nombrados previamente y con el tiempo da problemas ya que cuando existe el evento del sombreado la generación del sistema no cubre la carga del lugar o en un caso contrario al tener una excesiva irradiación no tener cómo almacenar para ser utilizada al ser necesaria, mediante este suceso la presente investigación basa su estudio en un algoritmo que ayuda a predecir el comportamiento de la irradiación solar para ser aprovechada en su totalidad [4].

Como se mencionó anteriormente la irradiación solar llega de una manera variable a la tierra esto ocurre que no sea predecible al momento de realizar una modelación de sistemas fotovoltaicos, esto produce en diseñadores un mal dimensionamiento en los equipos a utilizar para el sistema, para contrarrestar este contingente se analiza los algoritmos de predicción basados en redes neuronales que son parte de la inteligencia artificial [5].

En Inteligencia artificial su principal rama o subconjunto para trabajar en predicción de datos es Machine Learning, que mediante un aprendizaje supervisado usa un algoritmo interno de red neuronal para los datos de entrada irlos clasificando en clústeres y entrenándoles para tener una salida o una respuesta deseada, similar a la ingresada [19].

La presente simulación se elaboró en el software Python (Colaboratory), herramienta computacional que tiene un sin número de usos, pero en este caso fue elegido por las librerías que posee entre ellas una de las más importante la Inteligencia Artificial que en la actualidad está siendo muy empleada en la parte de energías fotovoltaicas que es el área que se necesitó para este proyecto.

**Planteamiento del problema:**

La hacienda “La Campiña” ubicada en la parroquia Mulaló, no consta con energía eléctrica por la distancia a la que se encuentra y el gasto que produce transportarla desde el ramal más cercano, por este motivo se optado por otro tipo de energía, pero para ser implementada se debe hacer un previo estudio que garantice la viabilidad de esta. En la actualidad no existe un estudio de irradiación solar en la parroquia Mulaló que permita realizar la predicción de generación fotovoltaica basado en aplicación de Machine Learning.

Machine Learning ayudará a predecir la falta o exceso de irradiación solar empleando un algoritmo de predicción basado en “Random Forests” y datos previamente obtenidos de medición en el lugar. Al no realizar un estudio previo no se conocerá la factibilidad para la implementación de este tipo de energías en la Hacienda “La Campiña” parroquia Mulaló, esto conlleva a no cubrir la demanda que es necesaria. Por ende, el usuario, deberá implementar elementos adicionales de generación térmica que representa mayor gasto, o cuando exista una excesiva de potencial de irradiación solar no tener el número de baterías adecuadas para almacenarla y utilizarle cuando sea necesario.

**Formulación del problema:**

¿Qué tan rentable es la aplicación de energías fotovoltaicas para cubrir la demanda diaria de la Hacienda “La Campiña”, mediante el análisis de irradiación solar?

### Objetivo General:

Analizar el potencial de irradiación solar mediante la aplicación de un algoritmo de predicción que permita conocer el comportamiento de la generación fotovoltaica y comparar con la variable real.

### Objetivos Específicos:

- Realizar una investigación bibliográfica para determinar conceptos que van a ser aplicados y de ayuda en el estudio.
- Adquirir datos del potencial de irradiación solar en un lapso de tiempo para análisis del comportamiento variable.
- Elaborar un algoritmo de predicción para el análisis de irradiación solar mediante la aplicación de Machine Learning y su modelo predictivo “Árboles de decisión”.
- Realizar la validación del método aplicado mediante análisis comparativo.

### Sistemas de tareas en relación a los objetivos específicos:

**Tabla 1:** Sistemas de tareas en relación a los objetivos específicos:

Objetivos específicos	Actividad (tareas)	Resultado de la actividad	Descripción de la actividad (técnicas e instrumentos)
1	Investigación Bibliográfica	Base de fichas bibliográficas.	Técnica del fichaje
2	Recopilación de datos	Potencial de irradiación solar.	Técnica de campo.
3	Elaborar un algoritmo de predicción	Machine Learning aplicado a la generación fotovoltaica.	Técnica experimental.

4	Realizar la validación del método aplicativo	Optimización del algoritmo.	Técnica experimental.
---	--	-----------------------------	-----------------------

## **Justificación**

En la actualidad la energía renovable se está aplicando con mayor frecuencia ya que es más conveniente su aplicación en lugares muy distantes donde no se accede a energía eléctrica o es muy costosa llevarla. Este es el caso de la Hacienda "La Campiña" ubicada en la parroquia Mulaló perteneciente al Cantón Latacunga, Provincia de Cotopaxi que para tener energía la red eléctrica se encuentra muy distante y esto conlleva a realizar una gran inversión en ser trasladada, la manera más factible para tener energía eléctrica es utilizando energías alternativas como la energía fotovoltaica. Pero para implementar un sistema fotovoltaico se debe realizar un previo estudio de campo sobre optimización y factibilidad.

El presente proyecto se elaboró con una previa obtención de datos en la Hacienda "La Campiña" ubicada en la parroquia Mulaló, con el objetivo de elaborar un algoritmo que ayude a optimizar datos y predecir el potencial de la irradiación solar con la aplicación de Machine Learning y la técnica estadística llamada "Árboles de decisión". Con este proyecto se pretende realizar un análisis del potencial de radiación solar existente en el lugar mediante un algoritmo de predicción esto ayudará a establecer estudios superiores de rediseño fotovoltaico, selección y optimización.

Además, con el presente estudio tendrá una base de datos reales, estos datos beneficiarán a diseñadores para posteriores implementaciones de sistemas fotovoltaicos, seleccionando los materiales y elementos de manera óptima que cubran el potencial de la carga requerida.

## **Hipótesis**

Analizando la irradiación solar, temperatura y la demanda por meses en La Hacienda "La Campiña", se crea un algoritmo de predicción, basado en Machine Learning utilizando una de sus principales modelos de predicción como Árboles de decisión, con dicho modelo podemos conocer si es viable o no la implementación de un sistema fotovoltaico.



# CAPÍTULO I

## FUNDAMENTACIÓN TEÓRICA- METODOLÓGICA

### 1.1 Antecedentes de la investigación o fundamentación del estado del arte.

En el actual capítulo, se visualizan antecedentes teóricos necesarios en la investigación que ayudan a llegar a los resultados deseados. El principal tema que se va a tratar es algoritmos de predicción en sistemas fotovoltaicos, estos temas se los analiza considerando la irradiación solar.

Mediante el estudio de los recursos renovables, la facilidad de su implementación y los beneficios que brindan el presente proyecto, basa su estudio en irradiación solar como materia prima para proyectos fotovoltaicos, como se conoce la irradiación llega de una manera continua al planeta, pero no es aprovechada adecuadamente, por este motivo se elaboró un algoritmo que me ayude a predecir y optimizar la irradiación solar, mediante una base de datos previa obtenida en el lugar del estudio. Las energías renovables son puras no producen daño al medio ambiente.

Desde la antigüedad la energía solar fue aprovechada de una manera básica como fuente de calor y luz para realizar construcciones. El autor [6] menciona que en años anteriores las energías renovables han sido la solución para los efectos del calentamiento global, el uso de combustibles fósiles que no son renovables y producen dióxido de carbono ( $CO_2$ ) esta emisión es la causante de la elevación de temperatura promedio anual en la atmósfera. Las plantas de generación a partir de energías renovables son competitivas con respecto a las plantas cuya materia prima son los recursos fósiles [7]. Mediante el estudio que se ha venido realizando para introducir las energías renovables a los sistemas convencionales de generación para el beneficio de la población y que lleguen a reemplazarlos debido a los efectos negativos mencionados anteriormente.

El cerebro es el órgano más importante del ser humano mediante el logramos apreciar imágenes, sonidos, olores y sabores. En él se guarda un sin número de recuerdos, emociones y permite soñar. Con esta visión el ser humano con el pasar del tiempo ha buscado y desarrollado máquinas que cumplan actividades similares, e incluso programas en softwares que ayudan a anticiparnos a un acontecimiento que vaya a suceder empleando Inteligencia Artificial (IA). No se puede definir adecuadamente a la IA pero los autores [8] argumentan que es la capacidad que tienen un individuo para actuar de una manera ordenada, la IA se divide en varios subramas de estudio entre ellas se encuentra Machine Learning que es la empleada en el presente estudio junto con algoritmos de optimización.

Mediante el software Python creado en el centro para las matemáticas y la informática por Guido Van Rossum en los Países Bajos con el objetivo de generar ideas y para ser empleadas por programadores en diversos campos comerciales, sociales y proporciona resultados completos y deseables [9], este Software fue elegido para llegar a los resultados planteados mediante una de sus principales librerías como Machine Learning que, con datos reales obtenidos anteriormente y un modelo de predicción arroja datos similares a los ingresados para que en una futura implementación del sistema fotovoltaico la demanda del lugar sea solventada en su totalidad.

## **1.2 Fundamentación Teórica**

### **1.2.1 Sistemas fotovoltaicos**

La tecnología solar fotovoltaica (FV) radica en la conversión directa de la radiación solar en electricidad, esta conversión es realizada por una celda solar cuya unidad básica es efecto fotovoltaico. La energía solar fotovoltaica está empleada en una amplia gama de aplicaciones donde se necesita producir electricidad para satisfacer las necesidades energéticas de quienes no tienen red eléctrica (sistema fotovoltaico autónomo) o para generar energía para la red eléctrica (sistema fotovoltaico conectado a la red). Los sistemas fotovoltaicos son una solución factible a lugares remotos, el montaje de sistemas fotovoltaicos domiciliarios o en pueblos, ayudan a

dar una vida más cómoda a la población. La luz de origen solar consta de un conjunto de radiaciones electromagnéticas con alta frecuencia, que lleva el nombre de espectro luminoso. La intensidad de la radiación luminosa cambia si aumenta o disminuye su frecuencia. La Figura 1.1, indica el comportamiento del espectro luminoso. [1] [10] [11].

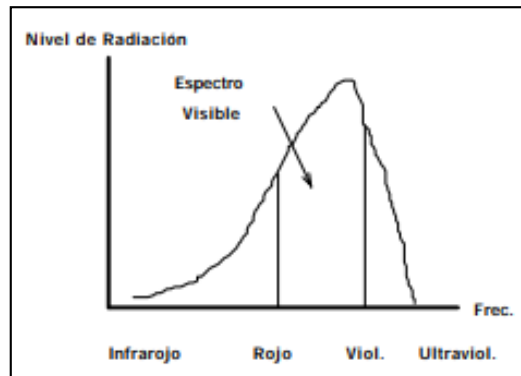


Figura 1.1. Espectro Luminoso [11].

Un sistema fotovoltaico está compuesto por diferentes equipos que cumplen una función específica:

- Un generador fotovoltaico.
- Una batería de acumulación.
- Un regulador de carga.
- Un inversor.
- El consumo.

#### 1.2.1.1 Generador fotovoltaico.

Transforma la energía del sol en energía eléctrica. Consta de la unión de varios módulos fotovoltaicos con una conexión serie y/o paralelo, tomando en cuenta que esta conexión se realiza de acuerdo a la necesidad si es serie la salida es de voltaje deseado, pero si es paralelo se obtiene corriente. Los módulos fotovoltaicos constan de células fotovoltaicas que producen electricidad con la captación de luz solar. La potencia que entrega un módulo depende del número de células que tiene, una célula puede suministra 3 W [6][10].

Las células solares están elaboradas con dos o más capas de semiconductores, las cuales forman un campo eléctrico fuerte para dividir cargas de signos distintos y generar energía eléctrica. Las células fotovoltaicas se encuentran dopadas para que su consistencia sea positiva y la parte negativa para dar lugar a la aparición del campo eléctrico, la más utilizada como elemento semiconductor es el silicio ya sea monocristalino, policristalino y amorfo como se muestra en la Figura 1.2, y en la Tabla 1.1, se detalla el porcentaje en condiciones ideales y rendimiento [6].



Figura 1.2. Celdas de Silicio Monocristalino, Policristalino y Amorfo [6].

Tabla 1.1. Rendimiento de paneles por su material de fabricación.

Tipo de Material	Rendimiento en paneles
Silicio Puro Monocristalino	16%
Silicio puro policristalino	14%
Silicio Amorfo	8%
Telurio de Cadmio	8%

La energía que emite un generador es almacenada si es necesario en un banco de baterías previamente

### **1.2.1.2 Batería Acumuladora.**

Cuando existe un exceso en la generación, es decir la radiación llega de una manera continua y con una alta frecuencia. Esta energía excesiva debe ser almacenada para ello se introduce en el modelo el banco de baterías donde la principal función es almacenar energía para ser utilizada en las noches que es donde más se necesita o cuando no exista suficiente radiación solar para cubrir la carga. La batería es el elemento que va a estar cargando y descargando constantemente. El elemento que controla la carga y descarga de la batería tiene el nombre de regulador de carga [10].

### **1.2.1.3 Regulador de carga.**

La batería ya que es un equipo que puede ser afectado por sobrecargas o sobredescargas excesivas, el regulador de carga es el encargado en proteger a la batería sobre estas contingencias que resultan dañinas y acortan su vida útil. El principio de funcionamiento se basa en: el regulador detecta que la batería empieza a sobrecargarse desconecta al generador FV, pero cuando llega a sobredescargarse, desconecta los consumos [10].

### **1.2.1.4 Inversor**

Son los más comunes llevan el nombre de acondicionamiento de potencia, aunque existen otras variedades que tienen diferentes aplicaciones como convertidores DC /DC, estos convertidores son empleados cuando se modifica la relación voltaje/corriente en una fuente DC. La Figura 1.3, muestra un sistema fotovoltaico con todos los elementos descritos [10].

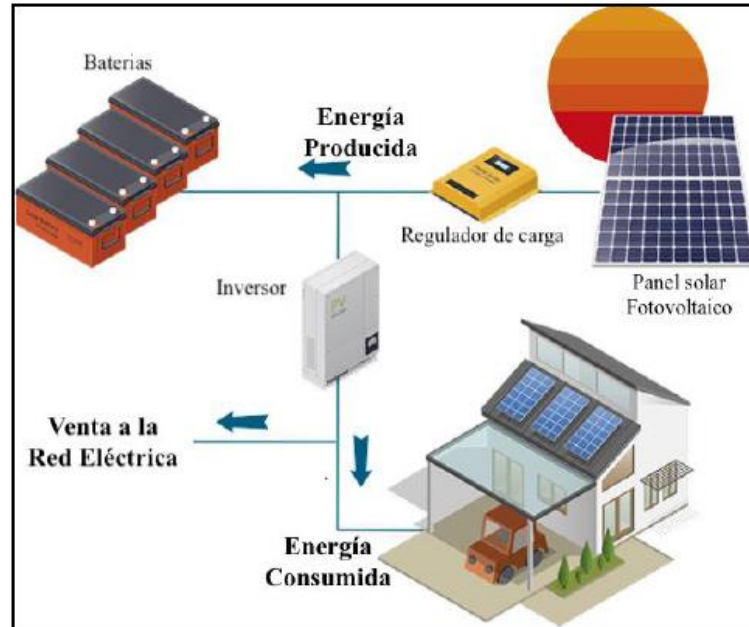


Figura 1.3. Bosquejo de un sistema solar fotovoltaico [2].

## 1.2.2 Clasificación de los Sistemas fotovoltaicos

La energía solar es una fuente inagotable no contaminante que produce electricidad mediante los sistemas fotovoltaicos, estos sistemas se clasifican en dos grandes grupos:

### 1.2.2.1 Sistemas fotovoltaicos Aislados de la red eléctrica (SFA)

Se denominan sistemas autónomos aquellos que no están conectados a la red eléctrica estos sistemas son los primeros que aparecieron y son más utilizados en lugares aislados. En la parte económica estos sistemas son muy rentables ya que su instalación no es muy costosa y el suministro es constante. Los sistemas autónomos son considerados desde sistemas simples, como un generador FV operando en DC, hasta sistemas que constan de un almacenamiento. En la Figura 1.4, se detalla las aplicaciones considerando 3 grupos. [10][12]

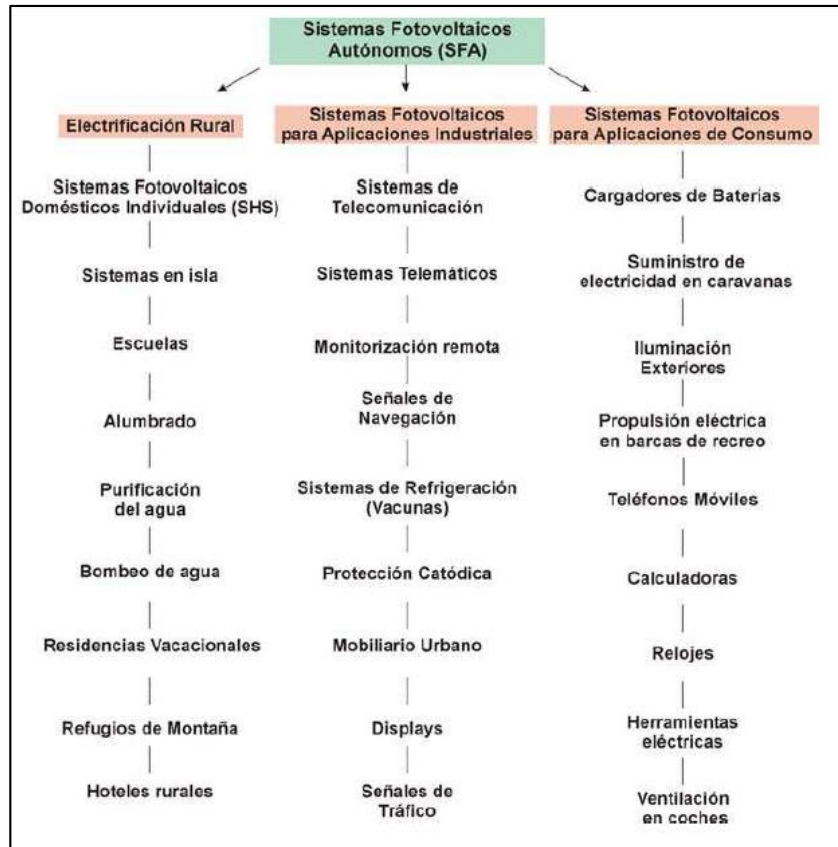


Figura 1.4. Sistema solar fotovoltaico autónomo [12].

### 1.2.2.2 Sistemas fotovoltaicos conectados a la red eléctrica (SFCR)

El objetivo principal de un sistema fotovoltaico conectado a la red eléctrica es aumentar la producción de energía eléctrica que es inyectada a la red. Mediante esta conexión un sistema pequeño residencial logra estabilizar sus parámetros de frecuencia y voltaje al instalar una carga. Los elementos principales del sistemas son los generadores fotovoltaicos, un inversor C.C. a C.A. y las protecciones seleccionadas para conectarse con la red, los elementos mencionados se aprecian en la Figura 1.5 [13].

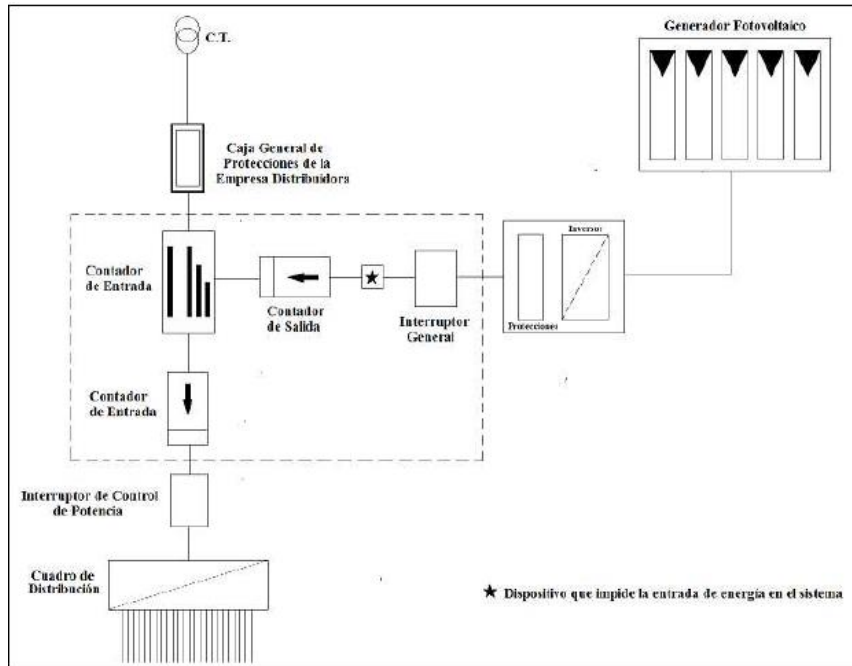


Figura 1.5. Esquema de un SFCR [13].

### 1.2.3 Radiación solar

La radiación se propaga mediante ondas electromagnéticas en todas las direcciones por el Sol en forma de radiación de onda corta con una longitud de  $0.29$  a  $0.25 \mu\text{m}$ , la radiación tiene es de gran importancia en áreas como: ingeniería, agricultura salud humana, entre otras, se emplea principalmente en la generación de electricidad al ser absorbida por un panel solar. La radiación solar tiene cambios dependiendo el horario, la latitud y las condiciones atmosféricas. Al pasar por el filtro de la atmósfera la radiación solar tiene cambios en su espectro [6][14].

La energía solar que llega a la tierra en 10 días al ser medida representa las reservas conocidas de petróleo, carbón y gas, en 86 minutos de radiación sobre la tierra representa un año de la demanda mundial. La radiación solar divide en 3 como son: directa, difusa y la suma de las dos que es la global, la radiación directa llega a la tierra con un valor de  $1\ 353 \text{ W/m}^2$ , que equivale a la constante solar y es medida con un instrumento llamado pirheliómetro, si es difusa en su trayectoria sufre afectaciones de fenómenos de reflexión y refracción, además de las nubes y elementos atmosféricos y se mide con pirómetro. La radiación global es la suma de las anteriores radiaciones como se indica en la ecuación (1.1) [6].



$$G = B + D + R \quad (1.1)$$

Donde:

G= Radiación global

B= Radiación Directa

D= Radiación Difusa

R= Albedo

La radiación Albedo o también considerada reflejada aparece cuando la emisión de radiación choca con un objeto elaborado de material de tipo espejo que rebota la radiación hacia el panel fotovoltaico, en la Figura 1.6, se indica el trayecto de la llegada de la radiación solar.

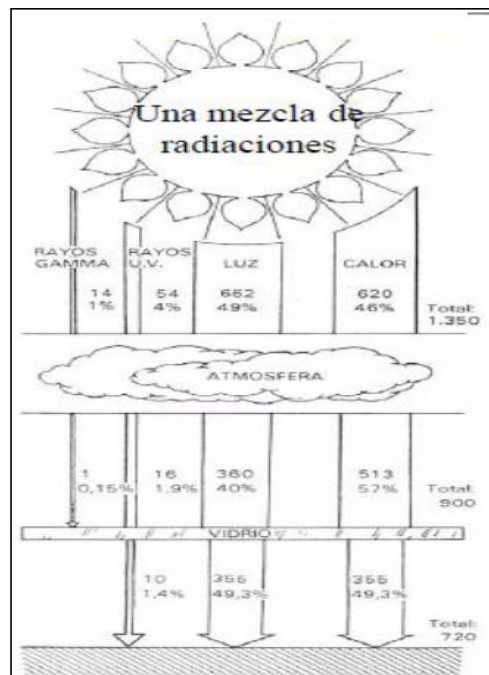


Figura 1.6. Radiación Solar [6].

#### 1.2.4 Irradiación Solar

La irradiación solar se define como la potencia de radiación solar por unidad de área incidente en una superficie la unidad de medida es el W, en un periodo temporal de irradiación la unidad de medida es Wh, en el sistema internacional [15].

$$I = G * \Delta t \quad (1.2)$$

Donde:

I= Irradiación Solar

G= Radiación global

$\Delta t$  = Período de tiempo

En la presente investigación se tomaron datos de irradiación solar para armar una base de datos para ser utilizada en el modelo de predicción, este proceso se elaboró con un piranómetro **Solar Power Meter SM206**, para la recolección de los mismos en un lapso de tiempo como se indica en la Figura 1.7 y Tabla 1.2.



Figura 1.7. Comprobación de datos de irradiación recolectados.

Tabla 1.2. Irradiación solar comprobada con paneles de diferente potencia.

<b>Irradiación solar</b>	<b>Irradiación solar para panel de 500w</b>	<b>Irradiación solar para 4000W</b>
100	13,1	183,8
200	45,0	630,0
400	108,8	1522,5
600	172,5	2415,0
800	236,3	3307,5
1000	300,0	4200,0
1100	331,9	4646,3

En el Ecuador existe un estudio previo de irradiación solar donde se han elaborado mapas en los que se detallan datos, estos mapas se encuentran en el Atlas solar, desarrollado por CONELEC, como se aprecia en la Figura 1.8.

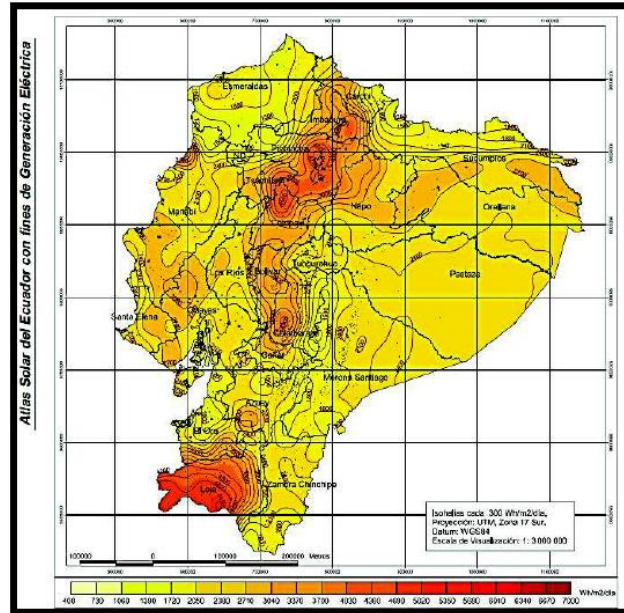


Figura 1.8. Atlas Solar Elaborado por CONELEC [16].

En la actualidad existen estaciones meteorológicas implementadas por instituciones gubernamentales que vienen realizando estudios sobre energía solar y plantean que puede ser aprovechada de una manera muy eficiente. La irradiación horizontal global y la irradiación normal directa son presentadas en las Figuras 1.9 y 1.10, respectivamente [17].

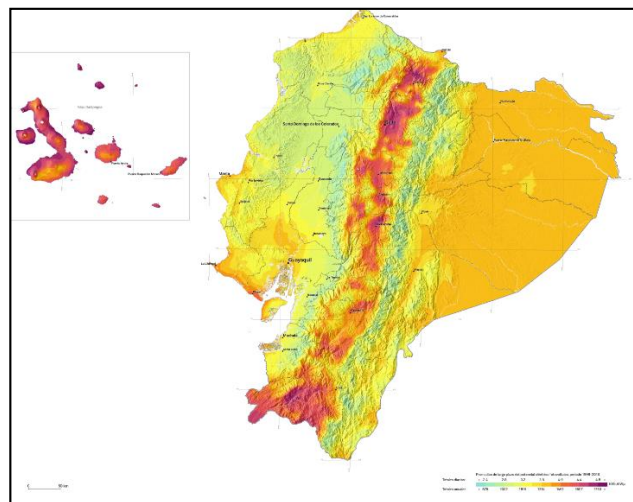


Figura 1.9. Irradiación horizontal Global Solargis.

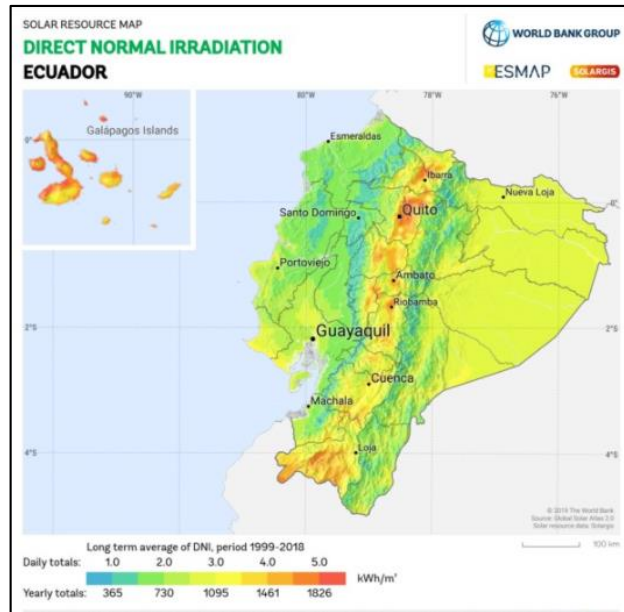


Figura 1.10. Irradiación Normal Directa Solargis.

### 1.2.5 Horas pico solares

El autor Frías en su trabajo define a las horas pico como la energía utilizable emitida por el sol obtenida por la integración de la irradiación respecto al tiempo, mencionado en horas de máxima radiación. Mientras que los autores Ramos y Luna define como la máxima absorción de energía solar por los paneles. Estas horas son de mayor intensidad de energía cuando más cercanos al Ecuador se encuentre. Para un conocimiento más explícito se plantea un ejemplo donde en un panel la hora pico marca 300 w/pico, es decir cuando el Sol irradie  $1\ 000\ W/m^2$  el panel daría 300 Wattios, para conocer la generación del panel se multiplica la HSP con la potencia del panel como se indica en la ecuación (1.3) [6] [14].

$$G_p = HSP * W_p \quad (1.3)$$

Donde:

$G_p$ = Generación del panel

HSP= Hora solar pico

$W_p$  = Potencia del panel

### Efectos de la emisión de ceniza volcánica en la irradiación solar.

La ceniza emitida por los volcanes tiene una composición de partículas de roca y minerales con una longitud menor a 2 mm, al ser emitida provoca un fenómeno conocido como “efecto refrigerante” que afecta a la irradiación solar ya que la absorbe y obstaculiza su paso a la superficie debido a que la ceniza se eleva a aproximadamente 12 km de la superficie. En la investigación presente el volcán que podría llegar a afectar a la hacienda es el Cotopaxi debido a que en los últimos años su actividad se ha incrementado y en años anteriores a existido alertas de supuesta erupción, la actividad del volcán en la actualidad es normal realiza emisiones de gases que no afectan directamente al paso de la irradiación solar. Pero si se llegara a dar reduciría en un 78 % la totalidad de irradiación solar captada y adicional a este análisis al empezar a caer la ceniza en donde exista energía a base de paneles solares estos no trabajarían a su totalidad su factor de planta se reduciría a un 15,5%. En la Figura 1.11. se muestra la caída en el porcentaje de irradiación solar en el planeta a lo largo del tiempo.

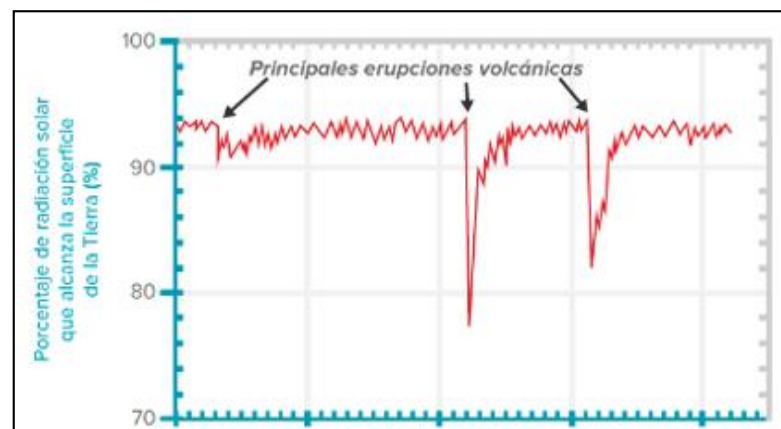


Figura 1.11. Porcentaje de caída de irradiación solar por emisión de ceniza.

### Ventajas del uso de Energía Fotovoltaica

- Para su ejecución no necesita combustible alguno ni genera ningún residuo tanto químico, electromagnético, ni térmico, lo que la convierte en no contaminante.

- La inversión para su instalación en la actualidad se ha ido reduciendo considerablemente, su vida útil es prolongada puede llegar a superar los 30 años.
- El mantenimiento es barato y sencillo, realizando una limpieza de la superficie y revisando conexiones y el cableado que tiene.
- La energía excedente se la puede almacenar para ser utilizada cuando sea necesaria mediante un banco de baterías.
- La principal ventaja es que se la puede utilizar en todo el planeta ya que la luz emitida por el sol llega hasta el último espacio del planeta y se considera inagotable y gratuita.
- Este tipo de energía es fácil instalación se puede ampliar, reducir el sistema o desmontarlo para ser trasladado, se puede instalar en el lugar de su necesidad.

### **Desventajas del uso de la Energía Fotovoltaica**

- La energía solar es variable es decir se recibe menos o más energía de acuerdo a la temporada.
- El no aprovechar toda la energía que llega a la tierra ya que los espacios donde se emplean son muy limitados.
- Se ve afectada por la contaminación atmosférica ya que la emisión de smog y polvo interfieren en la claridad de transmisión de la luz.
- La principal desventaja es la falta de conocimiento de la población y la falta de proyectos y estudio en este campo.

### **1.2.6 Inteligencia artificial**

La inteligencia artificial es una rama de la computación que apareció y comenzó a ser manipulada en el año de 1956 por un grupo de investigadores en la conferencia Dartmouth College, desde esta introducción en el campo de las ciencias se ha venido convirtiendo en el campo más relevante de investigación. Los autores Alasa y Araujo, define a la inteligencia artificial (IA) como la tecnología que se encarga de crear máquinas inteligentes que se asemejan al comportamiento del ser humano. La IA es la capacidad de las máquinas para desarrollar funciones asociadas a la

mente humana. Mientras que la autora Romero Rodríguez, define a la IA como la creación de Algoritmos o máquinas que imiten comportamientos humanos [18] [19].

Otra definición la IA, es un sistema informático entrenado para la toma de decisiones y medidas con la apreciación del entorno como, por ejemplo:

- Comprender situaciones, contextos.
- Llegar a identificar objetos y los significados.
- Analizar y resolver problemas.
- Elaborar tareas nuevas.
- Percibir el lenguaje natural.
- Reconocer imágenes.

Un ejemplo más claro es el ajedrez.

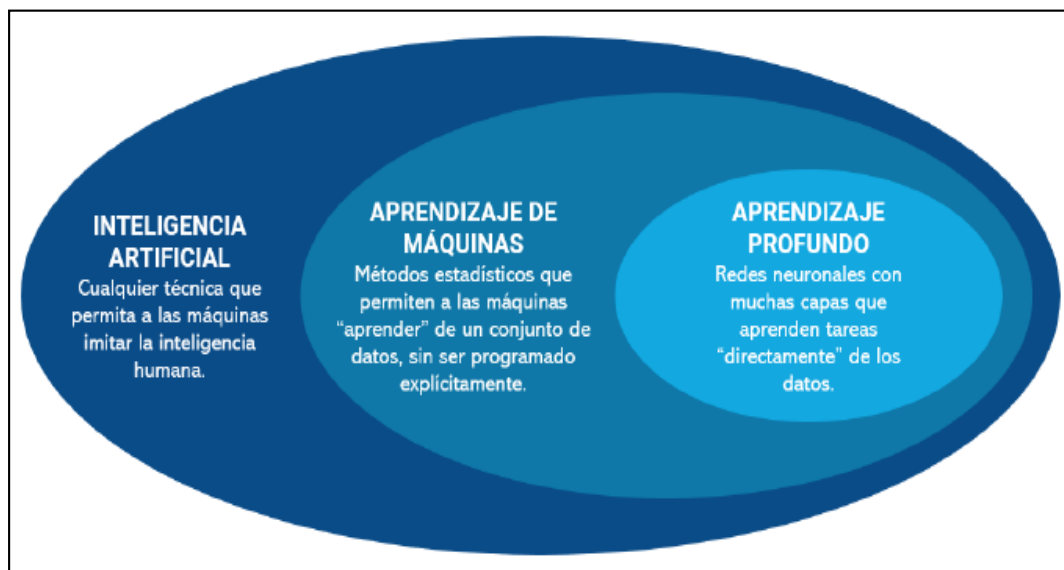


Figura 1.12. Diferencia entre la inteligencia artificial y los demás parámetros [19].

En la actualidad se ha tornado común los términos inteligencia artificial, aprendizaje de máquinas y aprendizaje profundo, pero estos tres tienen conceptos distintos como se observa en la Figura 1.12.

### 1.2.6.1 Aprendizaje de máquinas

El aprendizaje de máquinas es un tema que ya dentro de la inteligencia artificial, se define como una rama de la computación, que brinda a las computadoras la destreza para aprender sin ser claramente programadas. El aprendizaje de máquinas alcanza un sin número de algoritmos que se dividen en dos grandes grupos como son supervisados y no supervisados, estos dos grupos se derivan de etiquetas en los datos. Los datos se etiquetan cuando existen diferentes variables independientes y la variable dependiente de un modelo [19].

En la Figura 1.13, se detallan los dos grupos de aprendizaje mencionado.

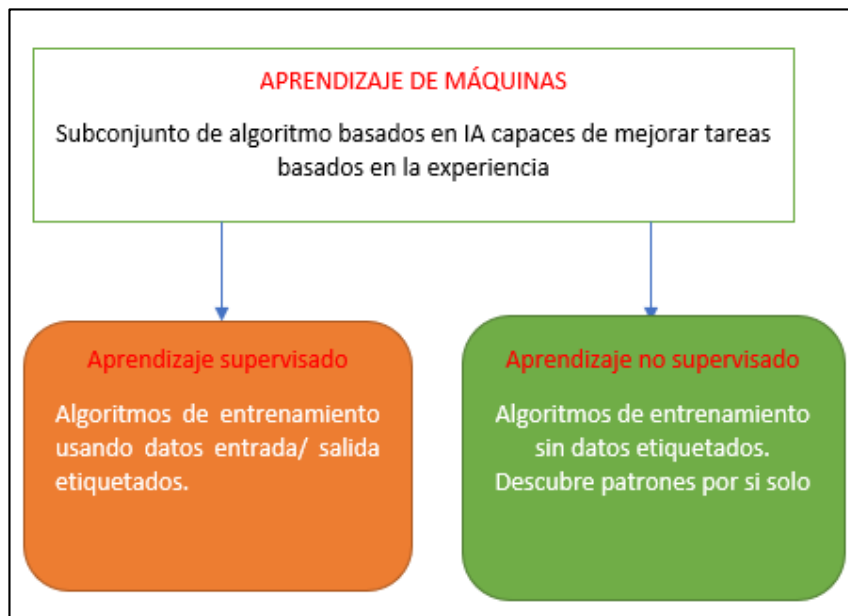


Figura 1.13. Aprendizaje de máquinas.

#### 1.2.6.1.1 Aprendizaje supervisado

El aprendizaje supervisado se define como una técnica de entrenamiento de un grupo de datos con el objetivo de reducir una función. Los datos de entrenamiento se deben etiquetar, es decir saber diferenciar cuáles son variables de entrada del modelo y las que son variables de salida. El aprendizaje supervisado es apto para clasificar datos, o para elaborar regresiones (predicción). Este aprendizaje emplea dos grupos de datos [19]:



1. El primer grupo entrega el modelo: halla los parámetros que simplifican el error.
2. El segundo grupo: se utiliza para clasificar la rentabilidad del modelo con medidas de error diferentes.

Los algoritmos de aprendizaje supervisado son métodos de aprendizaje automático que imitan el mecanismo de aprendizaje de seres humanos, se componen de neuronas que interactúan entre sí mediante diferentes niveles de capas. El proceso para llegar a una predicción se va realizando cuando a la neurona le llega información de la capa anterior y pasa a la siguiente capa, una neurona también se le puede denominar un modelo de regresión lineal múltiple ya que va internamente elabora una suma de la variable de entrada, al multiplicarlos por los pesos. Al final de la suma aparece una función de transferencia que introduce la linealidad de la neurona, y se conectan entre sí para formar una red. En la Figura 1.14, se aprecia una red neuronal de tres capas. La salida de una capa es la entrada de la capa continua [19].

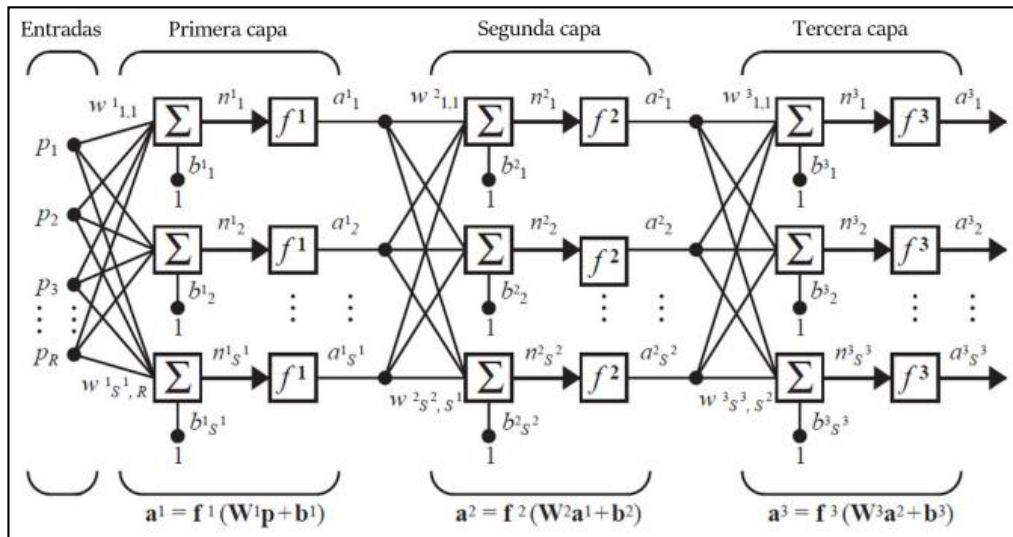


Figura 1.14. Red neuronal de tres capas [19].

$$a^3 = f^3(W^3 f^2(W^2 f^1(W^1 p + b^1) + b^2) + b^3) \quad (1.4)$$

Los pesos con los que se multiplican los valores de entrada toman el nombre de valores por ajustar. Los valores por ajustar en una red neuronal toman el nombre de

parámetros. El algoritmo de predicción más nombrado para encontrar el mejor valor de esos pesos es conocido como propagación hacia atrás [19].

### 1.2.6.1.2 Aprendizaje no supervisado

El aprendizaje no supervisado se caracteriza por tratar a todos los datos por igual ya que no tienen etiqueta como los supervisados. El algoritmo más utilizado es el K-means. Donde K es el número de grupos en los que se reparte la información. En un grupo los valores son parecidos entre sí, pero se diferencian de los demás valores de los grupos. En la Figura 1.15, se puede apreciar la función de un algoritmo de agrupación con datos bidimensionales [19].

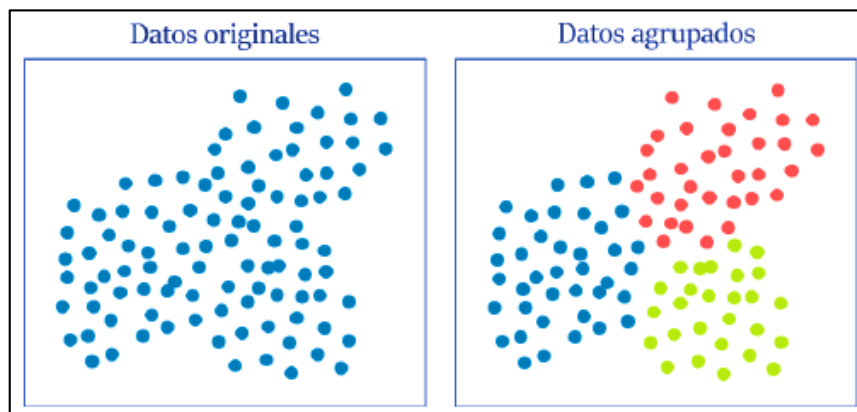


Figura 1.15. Algoritmo de agrupación con datos bidimensionales [19].

### 1.2.6.2 Machine Learning (Aprendizaje automático)

El autor Ferrer describe al Machine Learning como una rama de la inteligencia artificial. Tiene una aplicación amplia en las siguientes: clasificar textos, reconocimiento de voz, tareas de visión artificial para reconocer imágenes y detección de rostros, entre otros. Los autores Alasa y Araujo, define al Machine Learning como la capacidad de las máquinas para recibir y aprender de forma independiente un conjunto de datos, modificar y adaptar algoritmos, cómo procesan la información y aprenden sobre el entorno [18] [20].

El Machine Learning consta de muchos algoritmos cada uno con ventajas y desventajas, cada algoritmo no es mejor que el otro si no que cada uno funciona en el problema que se emplea [18].

Para utilizar un algoritmo se debe conocer diversos aspectos para diferenciar uno de los otros y emplear en la investigación el más apropiado:

- **Potencia y expresividad:** Los métodos de aprendizaje automático varían en riqueza y complejidad. Si la expresión es más pesada, se obtendrán resultados más precisos, pero existe el riesgo de transposición [18].
- **Interpretabilidad:** Los métodos poderosos a menudo crean modelos completamente consolidados, los algoritmos más simples son más fáciles de entender.
- **Usabilidad:** algunos métodos de Machine Learning aplican pocos parámetros o la toma de decisión es más simple para que sean fáciles de utilizar [18].
- **Rapidez de entrenamiento:** cada método actúa con una rapidez para ajustarse a los parámetros del modelo.
- **Rapidez de predicción:** trata de la velocidad de tomar una decisión de cada algoritmo.

Machine learning clasifica a los algoritmos en tres grupos:

1. **Supervised Learning:** Esto ocurre cuando el algoritmo aprende de los datos marcados y jerarquizados proporcionados por humanos. Al finalizar, las máquinas pueden generalizar y clasificar automáticamente sin ayuda por parte de nosotros, gracias al marcado. Gracias a este tipo de aprendizaje, las máquinas adquieren un conjunto de reglas que las “actualizan” y ayudan a tomar decisiones [18].

2. **Unsupervised Learning:** este aprendizaje no trabaja con datos etiquetados, ya que busca relacionarse por sí mismo. La máquina consta con la característica de detectar similitudes, diferencias o anomalías en las variables en base a la información [18].
3. **Reinforcement Learning:** el algoritmo aprende con la experiencia propia, es capaz de tomar la decisión más idónea en situaciones distintas [18].

Machine Learning se divide en dos tipos de aprendizaje supervisado y no supervisado a continuación, se dará a conocer la parte más importante de estos dos grupos y de los algoritmos correspondientes para luego profundizar en el modelo empleado en la presente investigación. Cabe recalcar que estos son utilizados en lo que se refiere a predicción de producción de energía solar.

### 1.2.6.3 Aprendizaje no supervisado en Machine Learning.

Mediante una minuciosa búsqueda bibliográfica se llegó a conocer los algoritmos que abarca el aprendizaje no supervisado, entre ellos el más popular es el algoritmo de agrupación K-means.

#### **Algoritmo de agrupación K-means**

La autora Romero Rodríguez, en su proyecto de investigación describe como parte del aprendizaje no supervisado es decir es necesario indicarle los datos de ingreso y salida para este algoritmo, la K viene a representar el número de conjuntos donde se agrupan los datos. El método busca colocar los pares de datos en el clúster que se encuentra más cercano, como se mencionó en un clúster los datos son similares. El algoritmo K-means es sencillo y robusto, se ejecuta con sencillos pasos para agrupar los datos de manera perfecta. Este proceso conlleva definir el número de clúster K, luego se coloca en los centroides los datos que se encuentran cercanos. Por último, recalcular los centroides con el promedio del clúster. Estos pasos se repiten hasta que los centroides no cambien de posición, de una manera más explícita se observa en la Figura 1.16. [19].

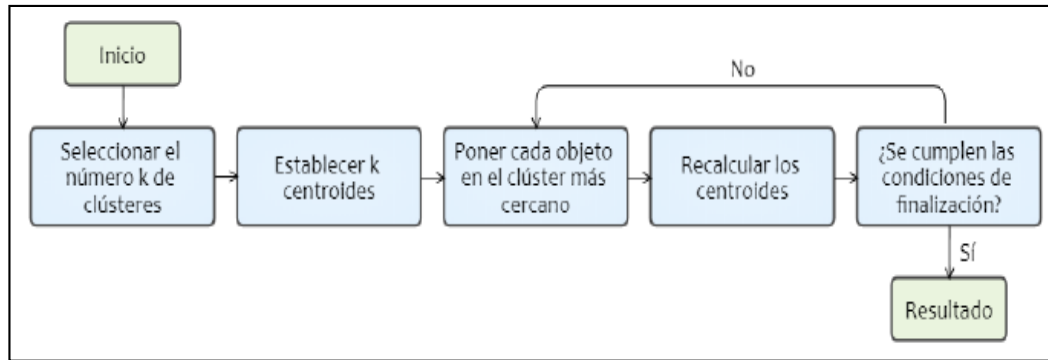


Figura 1.16. Diagrama de flujo con los pasos de K-means [19].

### El algoritmo en predicción solar

El algoritmo va generando los grupos de datos al entrenar todos los datos promedios por hora, mediante esto el algoritmo de agrupación entrega dos clústeres. En el primero se encuentra los datos de hora del día donde la potencia generada es más elevada, en segundo se van a colocar los datos de potencia más bajos esto sería la potencia de la noche y las horas de temprano en la mañana. Aparte de la división de datos donde son diferentes entre clúster y aparentes entre días, este algoritmo clasifica datos que tienen más alta irradiación correspondientes a días soleados, y también me clasifica en irradiación baja debido a días nublados poco soleados o escaso sol, se aprecia en el procedimiento el proceso que realiza el algoritmo K-means en diferenciar potencia generada y horas del día, utilizando un promedio de datos diarios.

Para la implementación de datos cercanos en los clúster más cercanos se emplea un proceso que realiza la medición de disimilitud entre dos puntos dados, utiliza el método de la distancia euclidiana entre dos puntos [19].

En la ecuación 1.5, se aprecia el método mencionado.

$$d_{euc} = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1.5)$$

Donde:

$d_{euc}$  = método de distancia euclidiana.

$n$  = número de observaciones.

$(X_i - Y_i)$  = son elementos cuya distancia es calculada.

En una manera resumida la agrupación K-means define los grupos para que la variación total sea mínima dentro de cada clúster. Esta variación también lleva el nombre de variación total dentro del clúster. Aunque el algoritmo estándar lleva el nombre de Hartigan-Wong, este algoritmo suma las distancias al cuadrado de las distancias euclidianas entre los elementos correspondientes como se describe la ecuación 1.6 [19].

$$WSS = \sum_{k=1}^k W(C_k) = \sum_{k=1}^k \sum_{x_i \in C_k} (X_i - Y_i)^2 \quad (1.6)$$

### **Método del codo**

El método del codo trabaja junto al algoritmo K-means, ya que en este al algoritmo se tiene una desventaja que es el de entregar un número de clúster como dato de entrada antes de conocer el mejor número de clústeres, y el método del codo provee del valor óptimo de K.

### **Método de silhouette**

Este método es adicional ya que verifica los resultados obtenidos por el primero. El método silhouette que mide la calidad de la agrupación. De manera corta comprueba que un dato tenga el parecido a los demás en el mismo clúster. Mientras más elevado sea el valor de este método más seguro es que el dato pertenece al clúster al cual fue asignado. Para cálculo del promedio se mide la distancia entre un punto con los demás puntos del clúster [19].

#### **1.2.6.4 Aprendizaje supervisado en Machine Learning**

El aprendizaje supervisado es el más empleado en predicción solar debido a su exactitud en predicción, la primera técnica en el aprendizaje es la clasificación. La clasificación se aplica cuando los datos son discretos. Los algoritmos donde se emplean los modelos de clasificación son [20]:

- Regresión lineal
- Árboles de decisiones.

- K-vecinos más cercanos.
- Ensemble.
- Máquinas de vectores.
- Redes neuronales.

#### **1.2.6.5 Regresión**

Considerando a la regresión como la segunda más importante después de la clasificación, se diferencia de esta porque los algoritmos de regresión trabajan con datos reales, se aplica para modelos de predicción. Al hablar de predicción solar, energía o cambio de temperatura es la más recomendada. La regresión plasma la idea en desarrollar datos de salida como una combinación lineal de los datos de entrada. Este método es utilizado en algoritmos como: árboles de decisión, Regresión de máquinas de vectores y redes neuronales [20].

#### **Regresión lineal.**

La regresión lineal es el método más simple se ha venido convierte en los principales algoritmos en el aprendizaje supervisado, este método interpreta la existencia de una dependencia lineal con las variables dependientes e independientes [18] [20].

#### **1.2.6.6 Árboles de Decisión**

Es el algoritmo más utilizado en problemas de clasificación. El árbol de decisión diseña modelos en forma de árbol, los cuales va dividiendo en subconjuntos de datos cada vez más pequeños, al mismo tiempo se va formando un árbol de decisión asociado. El modelo es compacto en datos ruidosos y aprende expresiones alternativas. El mecanismo actual divide los momentos organizándose en un árbol desde la raíz hasta llegar a un nodo hoja, con esto se llega a tener un árbol que tienen nodos de decisión y nodos hojas. El nodo hoja indica una decisión para un objetivo numérico. El nodo de decisión más elevado en el método árbol pertenece al mejor predictor llamado nodo raíz [18] [20].

En la Figura 1.17, se muestra la estructura que se plantea un árbol de decisión.

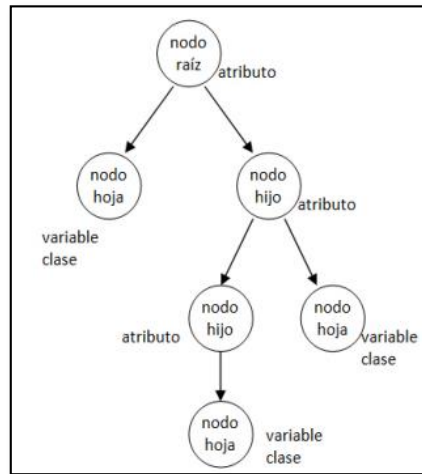


Figura 1.17. Estructura de un árbol de decisión [21].

### 1.2.6.7 K-vecinos más cercanos (KNM)

K-NN es un algoritmo que tiene muchas aplicaciones debido a su simplicidad y efectividad, este método no hace suposiciones sobre datos. Para entender este método se debe conocer el concepto de distancia, este método también considera la distancia entre dos puntos, existen diferentes estudios de distancias, pero la utilizada es la euclidiana. Este algoritmo elige K puntos en el conjunto de datos que consta de un patrón aparente al de la muestra (K vecinos más cercanos). Aparte de la función euclidiana ya nombrada existen dos más como la Manhattan y Minkawski [18] [20].

Estas funciones que me ayudan a predecir sistemas solares se ven con mayor claridad en las ecuaciones 1.7, 1.8 y 1.9 respectivamente.

### Ecuación Euclidiana

$$\sqrt{\sum_{i=1}^k (X_i - Y_i)^2} \tag{1.7}$$



### Ecuación Manhattan

$$\sum_{i=1}^k |X_i - Y_i| \quad (1.8)$$

### Ecuación Euclidiana

$$\left[ \sum_{i=1}^k (|X_i - Y_i|)^4 \right] \quad (1.9)$$

En la Figura 1.18, se muestra el proceso de este método.

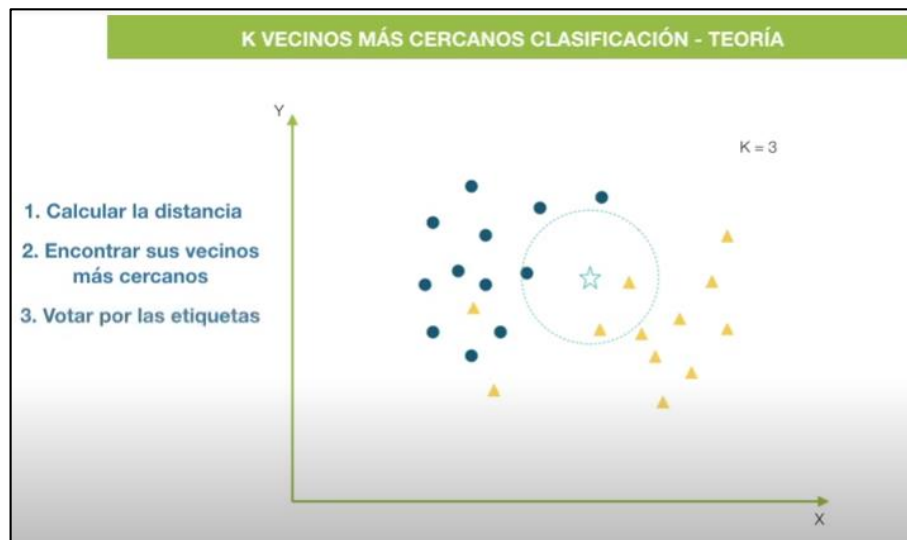


Figura 1.18. Teoría de K-vecinos más cercanos (Ligdi Gonzales).

#### 1.2.6.7.1 Ensamble

Se emplean bastantes algoritmos de aprendizaje en el presente método, para conseguir un rendimiento predictivo. El objetivo es combinar predicciones de varias estimaciones básicas con un algoritmo de aprendizaje, para mejorar la generalización y la robustez de la estimación. Hay dos grandes familias de métodos establecidos, que explicaré a continuación [18].

1. Averaging Methods: se caracteriza por crear estimadores independientes y promediar sus predicciones. La mezcla de los dos estimadores mejora el estimador base ya que produce la reducción de la varianza.

2. Boosting methods: Estos métodos generan estimaciones agregando gradualmente peso a los errores de entrenamiento anteriores.

### 1.2.6.8 Red Neuronal.

Las redes neuronales interpretan el trabajo de las neuronas humanas. Cada neurona recibe una señal de entrada y en base a la señal recibida envía una nueva señal o resultado a la red de salida. Una red neuronal se divide en capas en la cual la primera capa recepta la señal, mientras que la última capa es la salida que muestra los datos de clasificación. Las capas intermedias van colocando los pesos de acuerdo al resultado de la salida con el resultado deseado. La Figura 1.19 plantea la estructura de una red neuronal artificial [20] [22].

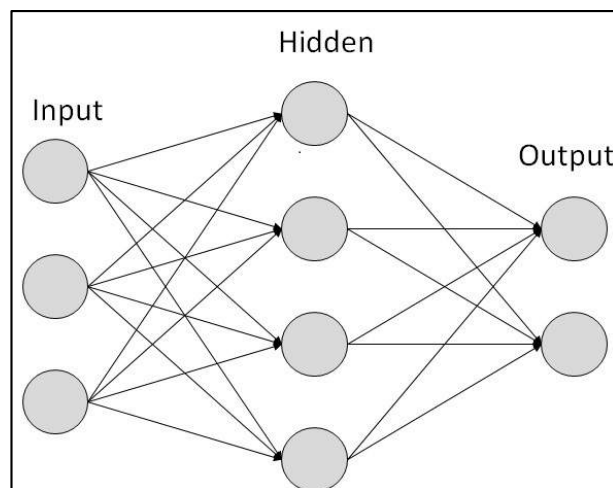


Figura 1.19. Red neuronal artificial [20].

### Red neuronal artificial (ANW)

En esta parte se practica la técnica de Machine Learning de redes neuronales artificiales. Utilizado en Serie temporal de red neuronal en la barra de herramientas Red neuronal. Esta aplicación será suficiente le permite resolver problemas de series de tiempo, como es el caso porque los predictores son datos históricamente. La ANW consta de tres problemas; el primero autorregresivo no lineal con entrada externa (NARX) este predice series  $y(t)$  dados  $d$  valores pasados de  $y(t)$  y otra serie  $x(t)$ , el segundo Nonlinear Autoregressive (NAR) predice series  $y(t)$  dados  $d$  valores caducos de  $y(t)$ , y el tercero, Nonlinear Input-Output predice series  $y(t)$  dados

valores pasado de serie  $x(t)$ , observar en la Figura 1.20, cabe recalcar que estos proceso no se desarrollan en la tesis [20] [23].

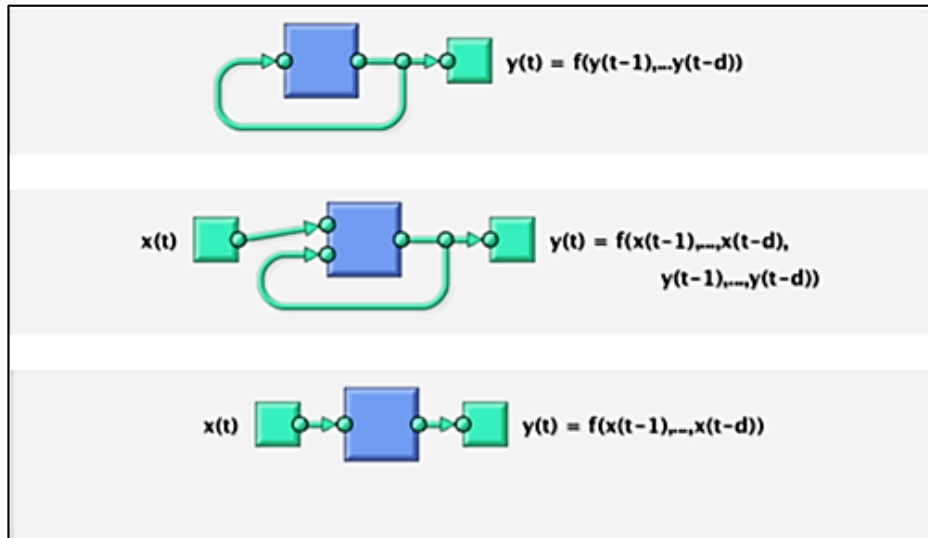


Figura 1.20. Red neuronal ANW [20].

### 1.2.6.9 Error en predicción

El error en la predicción no es más que la medición del rendimiento del modelo. Mediante el error se puede comparar entre diferentes modelos de los cuales siempre se utiliza el que presenta menos este error [19]. El error se mide con varias ecuaciones como se muestra en la Figura 1.21.

Siglas	Nombre	Fórmula
MSE	Error cuadrático medio.	$\frac{1}{N} \sum_{i=1}^N (y_{\text{pred}} - y_{\text{med}})^2$
RMSE	Raíz del error cuadrático medio o distancia media cuadrática mínima.	$\sqrt{\frac{1}{N} \sum_{i=1}^N (y_{\text{pred}} - y_{\text{med}})^2}$
nRMSE	Distancia media cuadrática mínima normalizada.	$\left( \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{\text{pred}} - y_{\text{med}})^2} \right) \cdot 100\% / y_{\text{med, máx}}$
MAE	Error absoluto medio.	$\frac{1}{N} \sum_{i=1}^N  y_{\text{pred}} - y_{\text{med}} $
MAPE	El error porcentual absoluto medio.	$\frac{1}{N} \sum_{i=1}^N \frac{ y_{\text{med}} - y_{\text{pred}} }{ y_{\text{med}} } \cdot 100\%$
MRE	Error medio relativo.	$\frac{1}{N} \sum_{i=1}^N \frac{y_{\text{pred}} - y_{\text{med}}}{y_{\text{total}}} \cdot 100\%$

Figura 1.21. Cálculo del error en predicción [19].

### 1.3 Fundamentación metodológica

#### 1.3.1 Tipo de investigación

El presente proyecto parte de una investigación bibliográfica para establecer conceptos de funcionamiento de sistemas fotovoltaicos, algoritmos de predicción y Machine Learning, con una base teórica solventada se llegó a una toma de datos actividad realizada en campo, con estos datos se crea un modelo de predicción.

La **Investigación bibliográfica** obtenida de varias fuentes como lo son tesis doctorales y artículos científicos, mediante esta investigación vamos solventando las dudas teóricas y conociendo cómo estructurar la base de datos, para obtener curvas donde se indica las horas de máxima irradiación solar y los meses con mayor recepción de irradiación solar.

La **investigación de campo** se desarrolló en la parroquia de Mulaló, en este lugar se ejecutarán los estudios. Se obtuvieron datos de irradiación solar, temperatura mediante un piranómetro para su análisis.

Mientras que con la **Investigación experimental** analizamos los datos obtenidos en el sitio mediante la ejecución de la simulación.

Con la **Investigación Aplicada** buscamos resolver el problema planteado, mediante el aprendizaje dado a lo largo de la ejecución del presente proyecto. Considerando la variación de la irradiación solar y la temperatura, estos datos fueron ingresados en una máquina de aprendizaje con el objetivo de una predicción del potencial solar a largo plazo.

#### 1.3.2 Método de investigación

La presente investigación basa su estudio en predecir irradiación solar mediante una modelación que emplea Machine Learning y un algoritmo de predicción basado en “Árbol de decisión”, mediante el software Python. La investigación parte de la obtención de datos de Irradiación solar y Temperatura variables principales para

poder plasmar el presente modelo de predicción, estas variables se ingresan al modelo elaborado, este proceso requiere estas variables para ir las compilando y entrenando de una manera idónea para brindarme una predicción similar a la real con un margen de error permisible que se encuentra dentro de las condiciones adecuadas.

### **1.3.3 Técnicas e instrumentos**

La presente investigación utiliza datos de irradiación solar y temperatura previamente obtenidos en el lugar donde existe esta problemática.

Para la recolección de datos de irradiación solar se ha utilizado un instrumento de medición llamado Piranómetro Solar Power Meter SM206, este instrumento me permite realizar la acumulación de la irradiación solar por un periodo de 5 minutos.

El software Python en la actualidad es de mayor uso para predicción ya que consta de un modelo de predicción llamado “Árboles de decisión”, este modelo es muy empleado para trabajar con variables de carácter estocástico.

## **Conclusiones del capítulo I**

- Con la profundización bibliográfica realizada en los diferentes algoritmos de predicción se logró definir el más adecuado para predecir la generación fotovoltaica.
- La investigación realizada en el presente capítulo da a conocer el desenvolvimiento y trabajo de la inteligencia artificial, su ramal principal como es Machine Learning y que tan útil es para realizar diferentes modelos de predicción.
- En la metodología se plasmó los pasos a seguir para llegar a obtener el modelamiento de predicción y la elaboración adecuada del modelo.

## CAPÍTULO II

### 2 PROPUESTA

#### 2.1 Título del proyecto

Análisis del potencial de irradiación solar en La Hacienda "La Campiña" de la parroquia de Mulaló. Propuesta para aplicación de un algoritmo de optimización basado en machine learning en la predicción de generación fotovoltaica.

#### 2.2 Objetivos del proyecto.

- Adquirir datos del potencial de irradiación solar en un periodo de tiempo para análisis del comportamiento variable.
- Elaborar un algoritmo de predicción para el análisis mediante la aplicación de Machine Learning y su modelo de predicción “Árboles de decisión”.
- Demostrar el procedimiento seguido en el software para el diseño del modelo de predicción considerando las variables utilizadas.

#### 2.3 Justificación

Para la implementación de un modelo de predicción de potencial fotovoltaica, utilizando árboles de decisión partiendo de la toma de datos de irradiación solar en el sector de la investigación, estudios iguales o similares no existen en el lugar o zona. La presente investigación parte de la selección del lugar, donde se colocó un equipo para la recolección de datos en un año, al tener estos datos reales se realiza una matriz de confusión para determinar que algoritmo de predicción me brinda una seguridad al momento de realizar una predicción, con este análisis se crea un algoritmo aplicando “árboles de decisión”, el software seleccionado fue (PYTHON Colaboratory) ya que en sus librerías consta de estas herramientas de predicción y su programación es de una manera más explícita.

## 2.4 Descripción de la metodología

La toma de datos se realizó en la hacienda “La campiña” de la parroquia rural Mulaló perteneciente al cantón Latacunga, con una Latitud de  $0^{\circ}46'40.33''$  S, Longitud  $78^{\circ}33'52.64''$  O y una Altura de 3 075 m, como se muestra en la Figura 2.1.



Figura 2.1. Lugar de la toma de datos. GOOGLE EARTH.

Hasta el día de la consulta el precio del kWh en Ecuador es de 0,1047 centavos, es considerada la más barata a nivel de otros países.

## 2.5 Metodología para cumplir los objetivos planteados.

Con una base de datos ya medida en el lugar de la investigación se empieza a trabajar para cumplir los objetivos planteados, estos datos fueron tomados en el lapso de un año, esta data se fue organizando y tabulando para obtener diferentes curvas promedio que se van a ir detallando en la presente investigación.

Los datos obtenidos con el piranómetro **Solar Power Meter SM206** se plasman a continuación, estos datos fueron tabulados en el software Excel para llegar a la curva de irradiación solar del mes de enero. En la Figura 2.2, se muestra la curva de irradiación de enero las demás curvas se encuentran en los ANEXOS desde el 1 al 22 respectivamente.



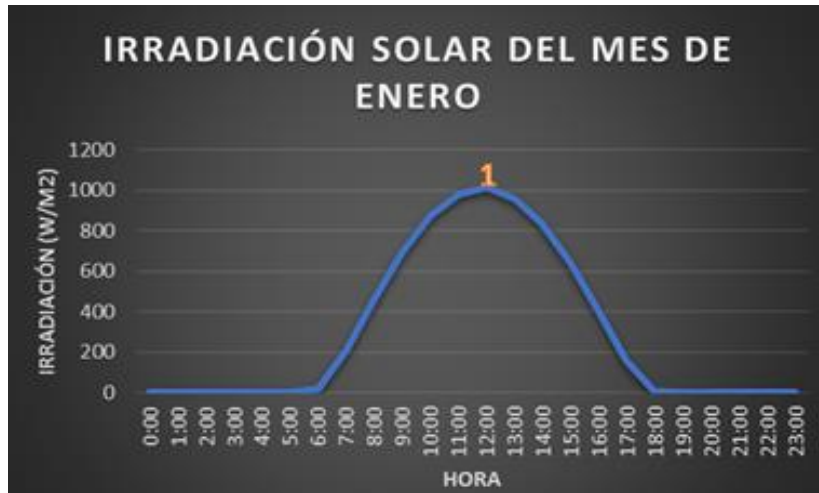


Figura 2.2. Curva de irradiación de enero.

Al analizar la curva del mes de enero se observa que la hora pico con mayor irradiación solar es a las 12:00 pm. En el punto 1; se tiene una lectura de  $998 \text{ w/m}^2$ . El instrumento comienza a detectar irradiación desde la 06:00 de la mañana donde va aumentando mientras va transcurriendo el día. Tomando en cuenta que mientras más despejado se encuentre mayor irradiación va a captar el instrumento.

En la Figura 2.3, se observa la curva de irradiación solar promedio de un día del mes de enero, con el objetivo de conocer a qué hora del día es más alta la irradiación solar. En el punto 1; empieza el análisis debido el instrumento de medición capta datos de irradiación solar arrojando un valor de  $22,38 \text{ W/m}^2$ , en el punto 2 se produce mayor irradiación solar con un valor de  $1\ 032,47 \text{ W/m}^2$  y en el punto de análisis 3 los datos de irradiación solar son bajas debido a que comienza las horas noche, dándome un valor de  $159,69 \text{ W/m}^2$ .



Figura 2.3. Curva de irradiación solar de un día del mes de enero

En la Figura 2.4, se observa la curva de variación de irradiación promedio del día del mes de enero los demás análisis de curvas se encuentran en ANEXOS del 1 al 22.

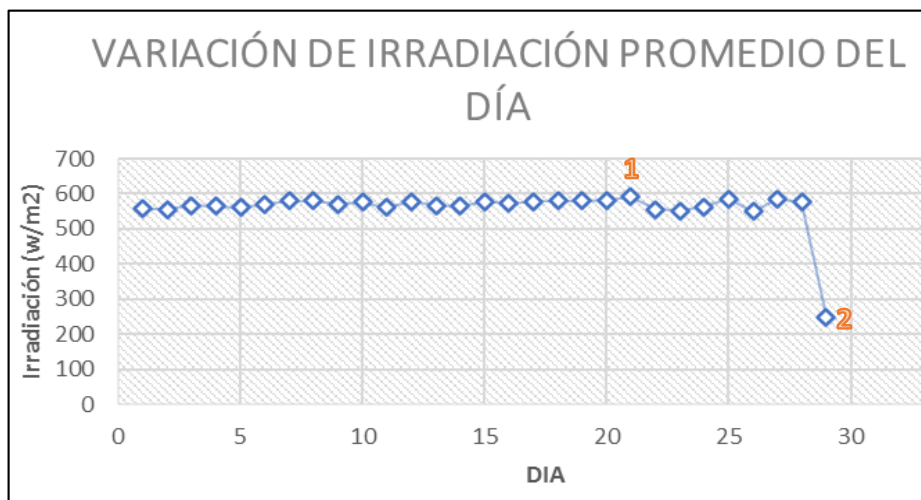


Figura 2.4. Curva de variación de irradiación promedio del día.

Obteniendo un promedio total del mes de enero podemos observar cómo va variando la irradiación cada día, donde el día más alto se indica en el punto 1, que llega a ser el 21 de enero con un valor de irradiación  $590,59 \text{ (W/m}^2\text{)}$ .

Los datos de irradiación solar de un año son plasmados en la Tabla 2.1. Estos datos son el promedio total de cada mes considerando que el análisis se realizó desde la

hora que existe irradiación es decir cuando el piranómetro comenzó a arrojar datos, hasta cuando ya no obtiene datos.

Tabla 2.1. Datos promedio anual de irradiación

MESES DEL AÑO	IRRADIACIÓN SOLAR
ENERO	605,7523
FEBRERO	613,4695
MARZO	623,7087
ABRIL	625,0447
MAYO	569,7847
JUNIO	554,3520
JULIO	561,1423
AGOSTO	598,8315
SEPTIEMBRE	629,5388
OCTUBRE	635,6378
NOVIEMBRE	612,9946
DICIEMBRE	558,3463

Con los datos de la Tabla 2.1, se realiza un diagrama de barras para el análisis de la irradiación del año en la Figura 2.5.

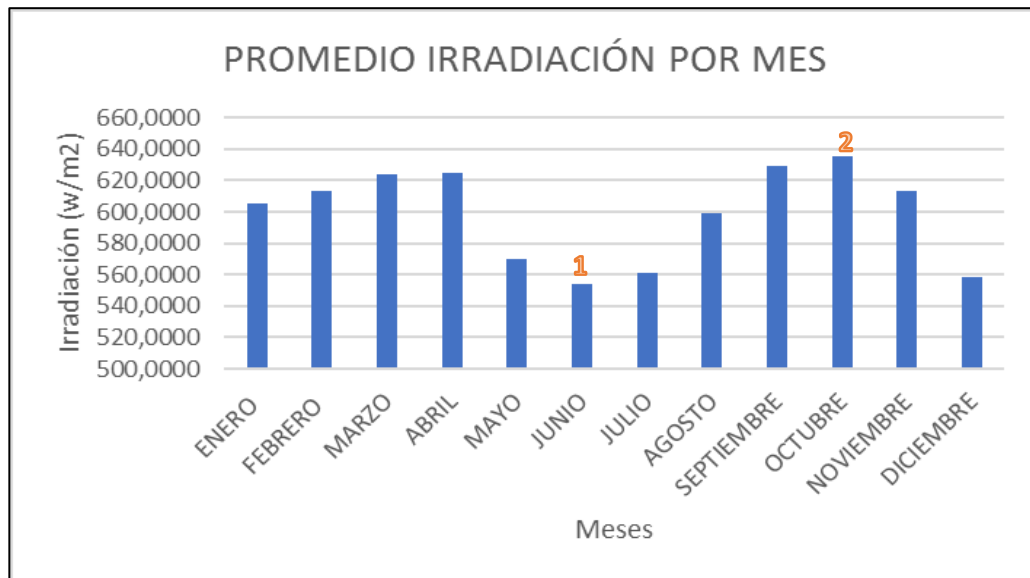


Figura 2.5. Promedio de irradiación por mes.

Con este diagrama de barras analizamos la variación de irradiación donde el mes más alto es octubre y el más bajo es junio. En la Figura 2.6, se observa la curva de irradiación solar en un año, dividida por sus respectivos meses.



Figura 2.6. Comportamiento de la irradiación solar promedio del año.

### 2.5.1 Histograma y frecuencia de irradiación solar

Para el análisis del histograma consideramos la tabla donde tenemos un promedio de irradiación por mes ubicada en el ANEXO 23 y la curva respectiva de la variación de irradiación diaria total en el ANEXO 24. Con el promedio total se tabularon 13 datos. La muestra se realiza con la ecuación (2.1) a continuación.

$$Muestra = \sqrt{\text{Promedio}} \quad (2.1)$$

$$Muestra = \sqrt{13}$$

$$Muestra = 3,6 \cong 4$$

En los datos tabulados se debe encontrar el valor máximo y mínimo para ingresar en la ecuación (2.2) para el rango y el ancho.

$$Rango = \text{máximo y mínimo} \quad (2.2)$$

$$Ancho = \frac{Rango}{Muestra} \quad (2.3)$$

$$Ancho = \frac{1006}{4}$$

$$Ancho = 251,52 \cong 252$$

Con el valor de ancho vamos armando nuestros intervalos para encontrar el histograma que se indica en la Figura 2.7.

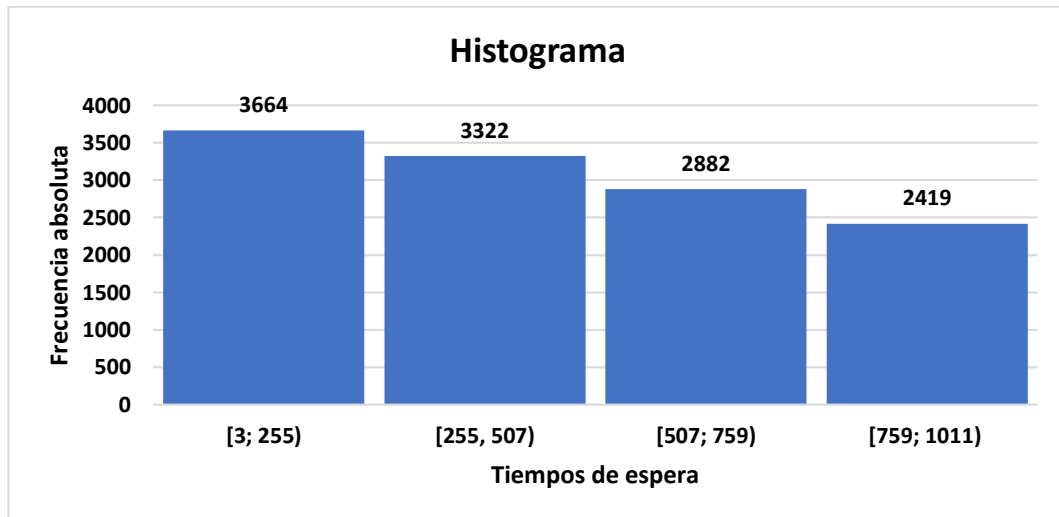


Figura 2.7. Histograma.

El histograma me identifica cómo se van distribuyendo las frecuencias absolutas, la tabla con frecuencia absoluta, frecuencia relativa, frecuencia porcentual se encuentran en el ANEXO 25.

### 2.5.2 Diagrama de Pareto de la irradiación solar

Los datos obtenidos previamente del histograma se van a ir ordenando adecuadamente de mayor a menor en un diagrama de Pareto, donde podemos observar el número mayor de repeticiones de frecuencia absoluta de 3 664 la de menor repetición de 2 419, obtenido también una frecuencia relativa acumulada de 0,59 como mínimo y un valor de 0,93 de máximo como se muestra en la Figura 2.8, en el ANEXO 26, se observa el diagrama de torta de las frecuencias.

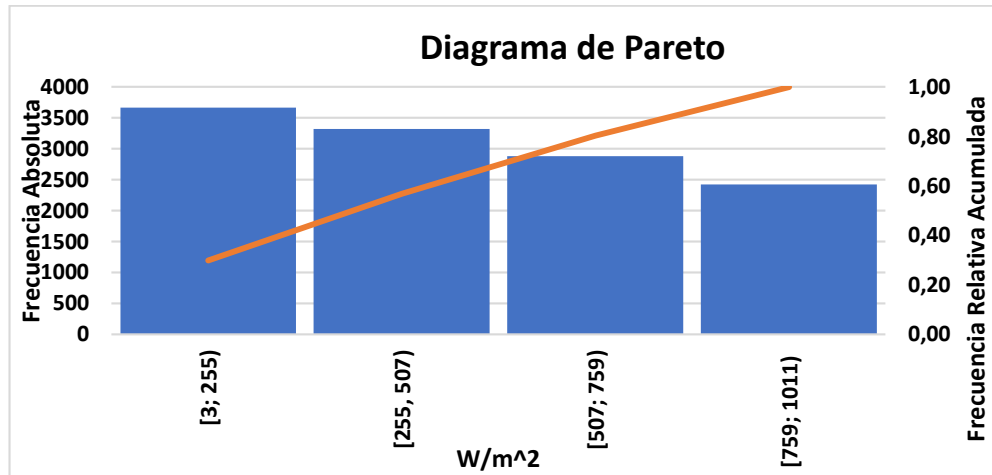


Figura 2.8. Diagrama de Pareto.

### 2.5.3 Análisis de la potencia requerida en el sector de la investigación

En el lugar donde se realizó el previo estudio como se mencionó en la actualidad no consta de energía eléctrica, es un lugar de descanso para turistas que buscan alejarse de lo cotidiano, pero en la hacienda se va a implementar electrodomésticos para aumentar el confort de las personas que visitan este lugar. Por ello se debe realizar un análisis de potencia requerida mediante la suma de potencias individuales instaladas para obtener una potencia total que necesita el lugar en la actualidad las 24 horas del día.

Tabla 2.2. Potencia de carga actual del sitio

HORARIOS	POTENCIA DE CAGA ACTUAL DEL SITIO								KW
	L	M	M	M	J	V	S	D	TOAL
1:00	0,350	0,421	0,286	0,302	0,421	0,380	0,520	0,480	0,395
2:00	0,350	0,200	0,400	0,350	0,158	0,370	0,718	0,350	0,362
3:00	0,250	0,124	0,124	0,254	0,321	0,458	0,421	0,368	0,290
4:00	0,425	0,254	0,365	0,540	0,325	0,540	0,354	0,425	0,404
5:00	0,360	0,210	0,660	0,460	0,654	0,645	0,458	0,456	0,488
6:00	0,354	0,481	0,654	0,541	0,325	0,396	0,568	0,654	0,497
7:00	0,875	0,892	0,587	0,478	0,894	0,965	0,896	0,987	0,822
8:00	1,254	0,987	1,250	1,680	1,230	1,125	1,365	1,254	1,268
9:00	1,120	1,350	1,254	1,540	0,980	1,540	1,480	1,365	1,329
10:00	1,300	1,540	1,845	1,750	1,650	1,870	1,654	1,541	1,644
11:00	1,335	1,354	1,458	1,254	1,365	1,845	1,630	1,870	1,514
12:00	1,654	1,245	1,360	1,365	1,360	1,250	1,650	1,870	1,469
13:00	1,365	1,354	1,258	1,654	1,250	1,147	1,365	1,680	1,384
14:00	1,250	1,125	1,368	1,654	1,548	1,654	1,354	1,870	1,478
15:00	1,250	1,365	1,540	1,365	1,258	1,980	1,540	1,840	1,517
16:00	1,230	0,840	0,900	1,250	1,654	1,750	1,620	1,640	1,361
17:00	0,250	1,650	1,254	1,215	1,250	1,360	1,540	1,540	1,257
18:00	1,250	1,450	1,250	1,200	1,124	1,025	1,540	1,600	1,305
19:00	0,758	1,020	0,954	1,110	1,250	1,458	1,540	1,035	1,141
20:00	1,025	1,320	1,020	1,105	1,254	1,235	1,245	1,412	1,202
21:00	1,125	1,365	1,254	1,369	1,654	1,254	1,664	1,524	1,401
22:00	1,254	1,654	1,250	1,245	1,548	1,624	1,254	1,254	1,385
23:00	0,878	0,548	0,587	0,870	0,780	0,715	1,650	0,982	0,876
0:00	0,758	0,254	0,354	0,458	0,745	0,845	1,112	1,020	0,693

La Tabla 2.2, muestra la potencia requerida en (kW) actualmente para la hacienda detallada en horas del día, en esta semana de análisis existe más consumo a las 15:00. En la Figura 2.9, se observa la curva de carga en (W), obtenida de la potencia total de la semana por horas.

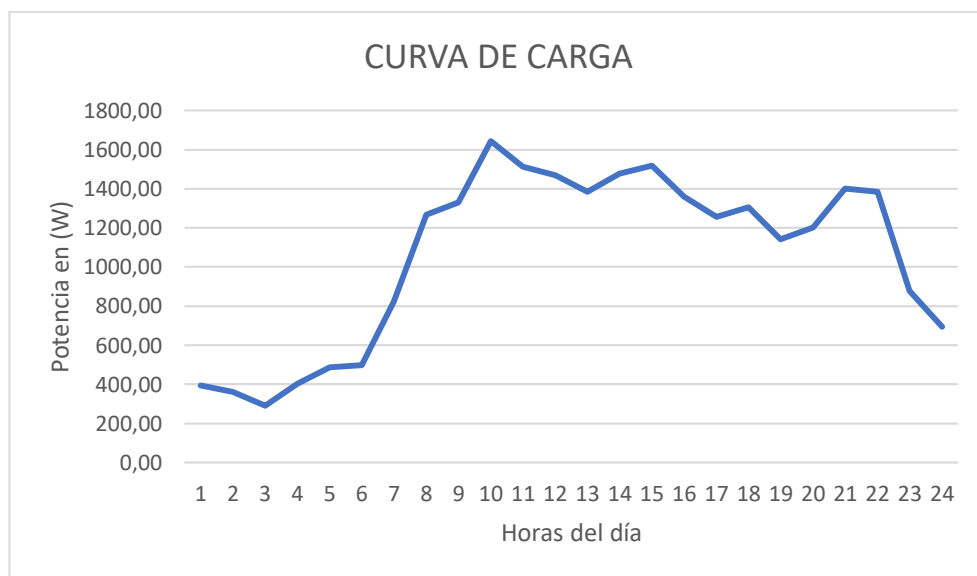


Figura 2.9. Curva de carga.

Para el cálculo de la potencia máxima generada por el panel solar ENERGYGRE ECOGREEN ENERGY de 5 00 W se aplica la fórmula donde se hace una relación de transformación como se muestra en la ecuación (2.4).

$$P_{max} = \left( \left( \frac{255}{800} \right) * irradiación \right) - \left( \frac{75}{4} \right) \quad (2.4)$$

Donde:

$P_{max}$  = Potencia máxima obtenida, se mide en (W).

irradiación = Es el dato de irradiación solar obtenido con un instrumento de medición se mide en (W)

$\left( \frac{255}{800} \right)$  = Relación de transformación del panel solar.

La ecuación debe ser empleada en cada medición debido a que la irradiación que se ingresa es de un periodo de tiempo específico es decir se aplica en la irradiación medida en un minuto, en un promedio de una hora de un día de una semana. En el capítulo III se aplica esta ecuación considerando una eficiencia de un inversor de 90 %. Cabe recalcar que este proceso se realiza para un diseño fotovoltaico específico, pero en la presente investigación se hace un breve análisis debido a que en el análisis de costos y rentabilidad se van a ocupar datos obtenidos de la ecuación (2.5).

$$Energía = P_{max} * Ef_{Iv}. \quad (2.5)$$

Donde:

Energía = Energía obtenida en un diseño fotovoltaico conectado a la red, es el total que va a ingresar a la red, se mide en (Wh).

$P_{max}$  = Potencia máxima obtenida, se mide en (W).

$Ef_{Iv}$  = Eficiencia del inversor.



#### 2.5.4 Factor de planta de la hacienda “La Campiña”

El factor de planta en la investigación es calculado con la demanda de la hacienda “La Campiña”, considerando que existe irradiación solar un periodo de tiempo por lo que el factor de planta va a ser menor o igual al 55 %, considerando que es afectado por días con poca radiación o por la nubosidad. Con la ecuación (2.6) calculó el Factor de Planta.

$$\text{Factor de planta} = \frac{\# \text{ horas operación}}{\# \text{ horas período}} \quad (2.1)$$

Donde:

# horas operación: Es las horas que el piranómetro arroja datos.

# horas período: Es el total de horas de un año.

Mediante los datos obtenidos y planteados en los ANEXOS 1 al 22 se relaciona y se llega a determinar que la generación provocada por la irradiación solar empieza a las 6:10 hasta las 18:00 en los días. Este periodo de tiempo se considera debido a que desde esta hora ya comienza el piranómetro a arrojar datos. En total tenemos 12h10min de sol. Asumiendo que el sistema que se puede llegar a implementar trabajaría a su máxima 7 horas, desde las 15:00 hasta las 23:00, se considera este horario debido a que la hacienda es un lugar de descanso rústico y en este horario llegan visitantes a descansar.

$$\text{Factor de planta} = \frac{8 \cdot 365}{8760} = 0,35$$

$$\text{Factor de planta} = 35\%$$

Mientras que si se calcula el factor de planta con las horas de sol al día se tendría.

$$\text{Factor de planta} = \frac{12,10 \cdot 365}{8760} = 0,504$$

$$\text{Factor de planta} = 50,4\%$$

Arrojando un valor del 50,4 % de Factor de Planta.

### 2.5.5 Matriz de confusión para selección del algoritmo

La matriz de confusión se aplica para conocer el algoritmo más eficiente al momento de realizar una predicción, mediante esta matriz podemos con un criterio decidir con que algoritmo voy a tener más exactitud y me va a brindar datos de predicción más cercanos a la variable real.

En la presente investigación se aplicó una comparación con tres diferentes algoritmos como lo es k-means, redes neuronales y por último árboles de decisión, siendo este el seleccionado debido a su precisión a comparación de los demás. La Tabla 2.3. indica el cálculo de precisión, exactitud y la función F1.

Tabla 2.3. Medidas de rendimiento

MEDIDAS DE RENDIMIENTO			
	K-MEANS	REDES NEURONALES	ÁRBOLES DE DECISIÓN
PRECISIÓN	0.72222	0.478021	0.478021
EXACTITUD	0.58333	0.479166	0.479166
F1	0.45054	0.448621	0.448621

El algoritmo árboles de decisión consta de una calidad en tareas de clasificación de 0.478021, una exactitud es decir una similitud al valor medido de 0.479166 y una función F1 de 0.448621, esta función se utiliza para conocer el rendimiento del modelo. En el Anexo 45 se plasman los datos obtenidos de la selección del modelo de predicción.

#### Matriz de confusión

Como resultado se llega a la Figura 2.10. donde se observa que la predicción al comparar datos de 200 y 500 se tuvo un acierto de 7 datos y una lectura errónea de 9 e 500 W, mientras que con 200 W brinda un acierto de predicción de 3 valores y una lectura errónea. Los resultados con los dos diferentes algoritmos se encuentran en el ANEXO 46 y 47 respectivamente.

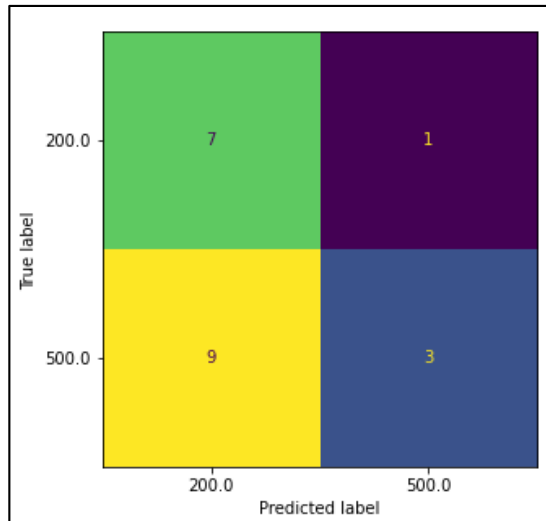


Figura 2.10. Matriz de confusión con árboles de decisión

### 2.5.6 Flujograma y pseudocódigo empleado en el análisis del proceso

La presente predicción emplea dos variables una la irradiación solar y la temperatura datos tomados de la estación en el lugar de estudio, mientras que la variable potencia que es la que va a ser objeto de predicción es asumida de acuerdo a la necesidad que se observa en el sitio. Estas variables van a ser ingresadas al programa de predicción elaborado. Esta predicción es elaborada con el algoritmo árbol de decisión que nos permite la predicción de generación fotovoltaica similar a la real, árbol de decisión hace predicción en base a datos obtenidos [24]. Se emplea el método de correlación que toma variables que más intervienen como irradiación, temperatura ambiente y potencia asumida.

El flujo de proceso que se muestra en la Figura 2.11, nos indica cómo va la estructura del programa partiendo desde el ingreso de librerías que van a ser utilizadas, entrenamiento de datos hasta llegar a una predicción similar a la real con un error permisible.

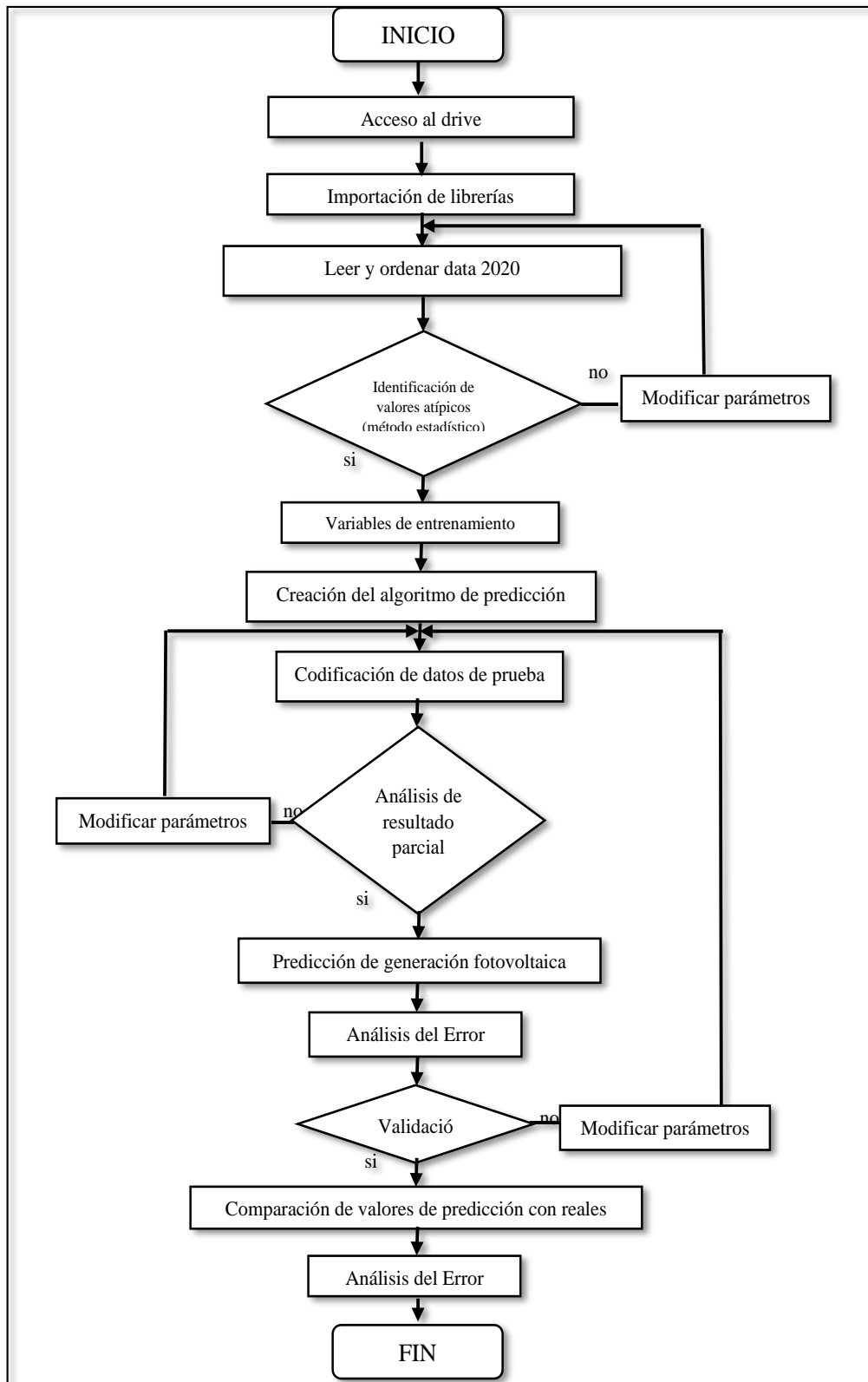


Figura 2.11. Diagrama de flujo del proceso

## **Preparación de datos**

Antes de ingresar los datos obtenidos por el instrumento de medición se comenzó a clasificarlos de una manera adecuada separando por meses y días, teniendo en cuenta el no alterar la concordancia de las dos variables utilizadas. Se obtuvo una base de datos completa de un año, la cual se fue tratando y discerniendo de una manera estadística. Para la ejecución del algoritmo se empleó las variables dos meses como enero, mayo y las de dos días de enero como el día 1 y el día 20.

## **Importación de librerías**

Numpy: para manipular matrices con números reales.

Pandas: librería que se utiliza para importar datos necesarios.

Matplotlib: librería empleada para gráficas.

Los archivos externos se cargan mediante la librería Pandas en un formato CSV, para la revisión en el algoritmo. Los archivos importados son los datos reales del día 1 de enero.

- `features = pd.read_csv('Datos_Solar_2020_dia1.csv')` # Importación de datos fotovoltaicos.
- `fecha = features.iloc[:,0]` # Fecha inicial para graficar.
- `potencia_inicial = features.iloc[:,8]` # Potencia original base de datos.

Para verificar que la matriz elaborada conste de datos válidos y no vaya a existir alguna incoherencia se emplea el comando `features.describe()`. La tabla de prueba se encuentra en el ANEXO 27 [25].

## **Análisis del conjunto de datos y manejo de datos faltantes**

Cuando se encuentran ya cargados los datos se verifica datos faltantes, datos que se encuentren mal cargados o en formato erróneo para tener una información útil y necesaria para ser procesado en las librerías y no me arroje error al momento de ser entrenadas. Los datos faltantes se completan con ceros [25].

### 2.5.7 Análisis de la división de DataFrame de entrenamiento, validación y prueba.

Para el funcionamiento adecuado del modelo se centra principalmente en su entrenamiento varios programadores dan un porcentaje del 80 % a este punto donde se va a ingresar datos ya obtenidos, para que vaya entrenando la neurona, teniendo presente que entre más datos se ingrese más exacta va a ser la predicción. Para este punto se ingresó la base de datos con las variables de irradiación solar, temperatura y potencia obtenidos en el sitio ya sea de un día de un mes o de un año, estos datos se muestran en la Figura 2.12. [25].

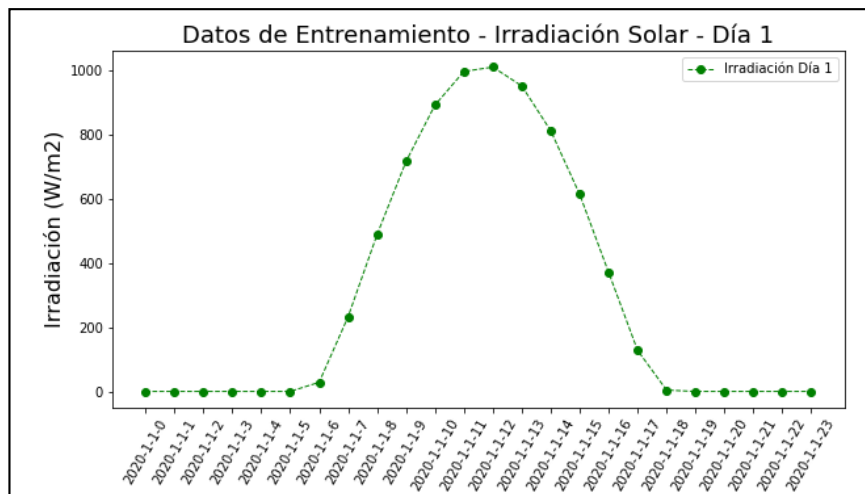


Figura 2.12. Datos de entrenamiento del Día 1 de enero.

La Figura 2.12, brindan una descripción clara de la temperatura obtenida el primero de enero estos datos se obtuvieron en grados Celsius.

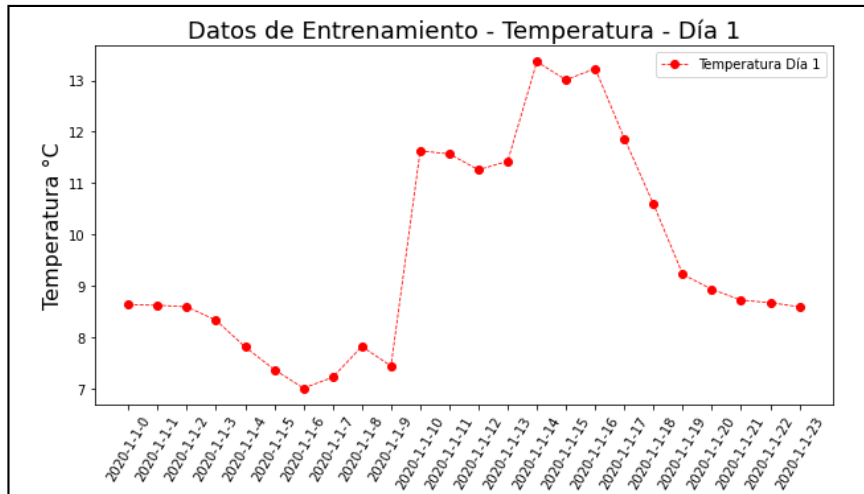


Figura 2.13. Temperatura de entrenamiento del Día 1 de enero.

El modelo de predicción consiste principalmente en eliminar la variable que se va a predecir, el modelo se va entrenando con las demás variables, al realizar una predicción del día 20 de enero se entrena al modelo con las variables del día 1 del mismo mes proceso plasmado en la Figura 2.14, con ello el algoritmo me arroja datos similares a los que fue entrenado con un margen de error permisible.

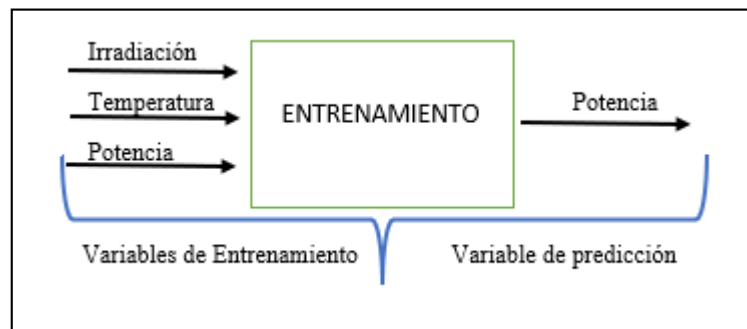


Figura 2.14. Proceso de entrenamiento.

### 2.5.8 Modelo de la Red Neuronal (Random Forest)

Machine Learning brinda una variedad de modelos para aprendizaje, pero para datos de registro que emplean secuencia se utiliza redes neuronales por la habilidad de almacenar en su memoria la información. La creación del modelo debe tener un orden para separar datos temporales una predicción por horas consta de una configuración donde tiene 1 entrada para diferentes horas y 1 salida por hora debido

a la configuración: Entrada 64 neuronas, capa oculta de 32 neuronas y salida 1 neurona, analizando el error de predicción [25].

El modelo árboles de decisión incluye una técnica de clasificación supervisada, la estructura consta de un árbol que se compone de raíz, nodos, ramas y hoja. Árbol de decisión tiene una base estructurada jerárquicamente donde la decisión final que se considera es que cumpla las condiciones planteadas desde la raíz hasta sus hojas, la estructura se ejecuta mediante los valores de las variables aplicadas [26].

La presente investigación utiliza random forest para analizar las variables de mayor incidencia en la generación de energía para que el predictor sea preciso y el tiempo de respuesta sea menor. Aparte con el estado del arte el algoritmo reacciona adecuadamente a esquemas con parámetros estocásticos.

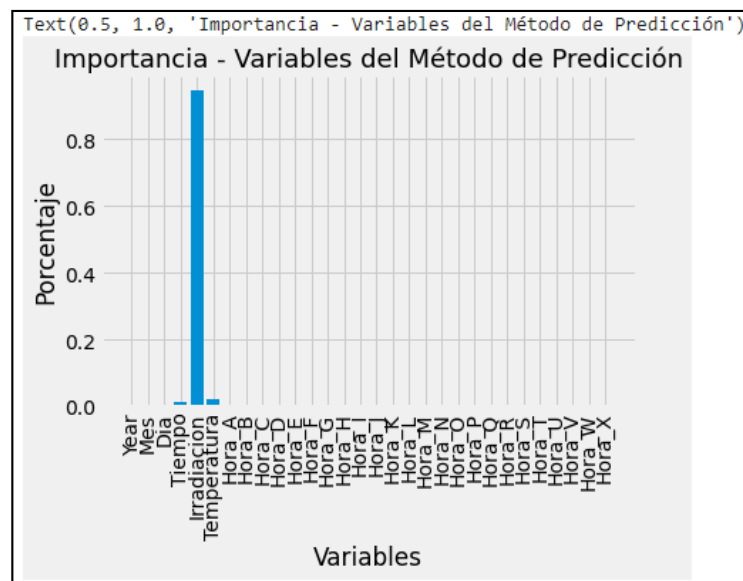


Figura 2.15. Variables con mayor incidencia.

Para el modelo la variable de mayor incidencia es la irradiación solar, a esta variable se le realizó un tratamiento exhaustivo, la variable temperatura al ir a la par con la irradiación también fue tratada y analizada, mientras que las demás variables empleadas para dar nombre y llenar valores nulos no son consideradas en este análisis porcentual, este análisis se plasma en la Figura 2.15.



## Árboles de decisión en aprendizaje automático

Aparte de ser un método estadístico un árbol de decisión ayuda a crear modelos predictivos automatizados, el método que utiliza se basa en utilizar variables independientes para predecir otra variable tomando en consideración el valor de la de la misma. Los nodos siempre van a representar datos a cambio de decisiones, este tipo de árbol se conoce como “árbol de clasificación” cada ramificación va a contener condiciones o reglas. Estas reglas tienen el nombre de “Reglas de decisión”, cada valor de las variables forma una cláusula de tal manera que si las variables ingresadas en este caso son: irradiación solar, temperatura y potencia de un día específico se compara con las variables del día que se va a predecir vamos a llegar a tener un resultado X que puede ser favorable o erróneo [27].

### Codificación de la matriz de datos

La codificación aplicada es para identificar variables y relacionarlas entre ellas, con este proceso no sólo se identifica las variables si no que se va creando un vínculo entre los datos de la investigación, con este proceso examinamos y estructuramos de una manera aplicada. Las variables codificadas se fueron ordenando para de una manera clara actuar sobre las necesidades de predicción, la línea de código utilizada para codificar se plasma en la Figura 2.16.

```
##### Proceso de codificación de la matriz de datos para el método de predicción
features20 = pd.get_dummies(features20) # Línea de código para codificación de la matriz de datos features.iloc[:,4]
features20.to_csv('M_Prueba_Codificada_dia20.csv') # Exportación matriz codificada para predicción
features20.iloc[:,5:].head(5) # Codificación de la matriz de datos para el método de predicción
print('The shape of our features is:', features20.shape) # Verificación tamaño de la matriz codificada
#####
Labels20 = np.array(features20['Potencia']) # Columna de datos originales para comparar
features20 = features20.drop(['Potencia'], axis=1) # Eliminar columna "Potencia" de la matriz de prueba
feature_list20 = list(features20.columns)
print('The shape of our features is:', features20.shape)
features20.to_csv('M_Prueba_Codificada_noFO_dia20.csv') # Exportación matriz de prueba sin la columna Potencia
features20 = np.array(features20) # Conversión de la matriz de datos a números reales para ingresar al método de predicción
# IMPORTANTE Features: Matriz de Entrenamiento
# IMPORTANTE Labels: Columna a predecir - Entrenamiento
# IMPORTANTE Features20: Matriz de prueba
# IMPORTANTE Labels20: Columna a predecir - Prueba
```

Figura 2.16. Proceso de codificación de datos o variables.

Donde los códigos de mayor importancia van a ser:

FEATURES: Que viene siendo la matriz de entreno.

LABELS: La columna que se va a predecir.

FEATURES20: Matriz de prueba.

LABELS20: Columna a prueba que se va a predecir.

### **Creación del modelo de predicción**

El modelo fue creado utilizando un estimado de 1000 árboles de decisión o Random Forest y 42 Random State, se utilizó estos parámetros debido a la cantidad de datos que se obtuvo considerando que un modelo de predicción entre más datos se vaya ingresando más se asemeja al real, la línea de código empleada va a ser la siguiente.

```
rf = RandomForestRegressor(n_estimators = 1000, random_state =  
42) # Instanciar el modelo con 1000 árboles de decisión
```

#### **2.5.9 Entrenamiento del modelo de predicción**

La parte principal del modelo es el entrenamiento, considerado por algunos autores que el entrenamiento es el 80 % del proceso de predicción, debido a esto se debe tener una base de datos completa para que la predicción se asemeje a la real es decir no me brinda un error amplio. Para el proceso de entrenamiento realizado se preparó al modelo con datos de las variables de irradiación solar, temperatura y potencia como se ha venido mencionando de un día para predecir la potencia de otro día.

Se tomó datos completos del 01 de enero para predecir datos de potencia del 20 de enero, el proceso se da de la siguiente forma, al modelo le ingreso las variables mencionadas del día 01 mientras que del día 20 no le indico la columna que deseo predecir en este caso la potencia los gráficos de irradiación y temperatura del día 20 se encuentran en el ANEXO 28 y 29 respectivamente.

```

from sklearn.ensemble import RandomForestRegressor # Librería de predicción por árboles de decisión
import pandas as pd
rf = RandomForestRegressor(n_estimators = 1000, random_state = 42) # Instanciar el modelo con 1000 árboles de decisión
rf.fit(features, labels); # Entrenamiento del método con los datos de entrenamiento fotovoltaicos del día 1
predictions = rf.predict(features20) # Predicción en base a los datos del día 20
potencia_pre = pd.DataFrame(predictions)
potencia_ori = pd.DataFrame(labels20)
M_Comp = {'Potencia_Original' : [potencia_ori],
          'Potencia_Predecida' : [potencia_pre]}

```

Figura 2.17. Proceso de codificación de datos o variables.

En la Figura 2.17, se indica las líneas de códigos empleados para la predicción donde se instaló el modelo con 1000 árboles de decisión, se entrenó el modelo fotovoltaico con datos del día 01 y se llega a una predicción base de datos del día 20.

### 2.5.9.1 Entrenamiento mensual del modelo

La presente investigación consta de una predicción mensual, con los datos obtenidos se entrena el modelo para la predicción de potencia de un mes. Esta predicción lleva la misma metodología que la predicción diaria, el modelo se entrenó con todas las variables de un mes, mientras que la variable del mes a predecir va a ser oculta no se le va a mostrar al modelo, solo se van a importar las demás variables para un entrenamiento adecuado.

Lo primero que se realizó es la clasificación de las variables obtenidas, mediante el pseudocódigo escrito se fueron dibujando las curvas de las variables más importantes como lo son irradiación solar, temperatura y potencia del mes 03, en este caso marzo. El pseudocódigo para las gráficas de curvas se encuentra en el ANEXO 30. Las curvas obtenidas se plasman en la Figura 2.18.

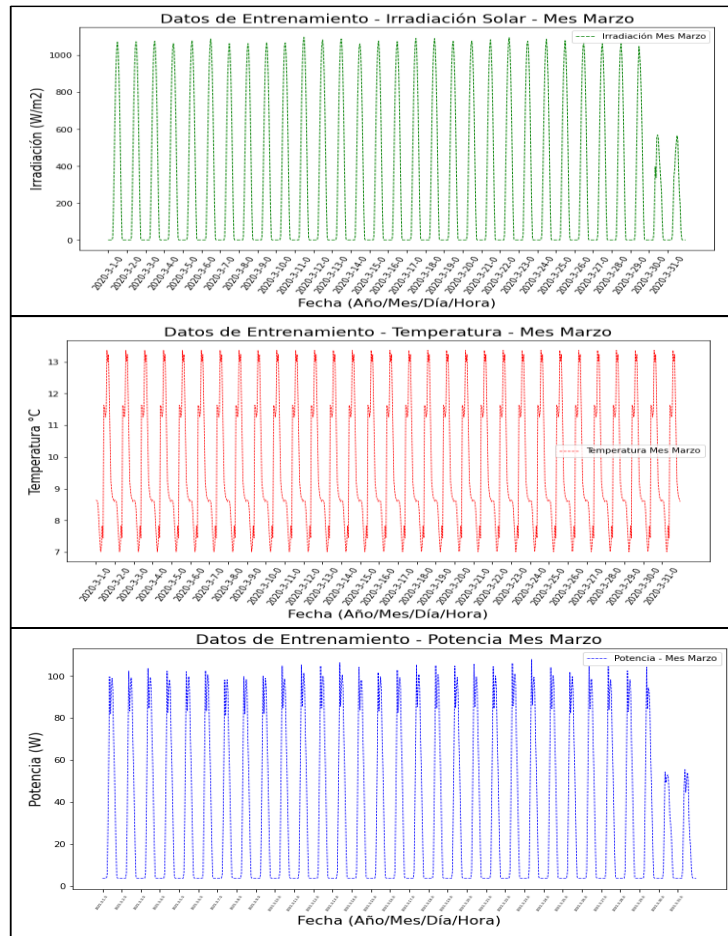


Figura 2.18. Curvas del Mes de Marzo

En la Figura 2.18, en la parte superior se observa la curva de Irradiación solar, en la parte media la curva de Temperatura y en la parte inferior la curva de potencia del mes de marzo, mes utilizado para el entrenamiento en el modelo de predicción elaborado con árboles de decisión.

### Proceso de codificación de la matriz como datos de entrenamiento

En este proceso se exportó la matriz codificada para el proceso de predicción, se verificó el tamaño de la misma, se oculta la columna de la variable que se va a predecir se vuelve a exportar la matriz, pero sin la columna de predicción y por último se cambia la matriz a números reales para ingresar al modelo, este proceso realizado en el software con sus líneas de código se muestra en la Figura 2.19.

```

##### Proceso de codificación de la matriz de datos como DATOS DE ENTRENAMIENTO
features3 = pd.get_dummies(features3) # Línea de código para codificación de la matriz de datos features.iloc[:,4]
features3.to_csv('M_Entrenamieto_Codificada_mes3.csv') # Exportación matriz codificada para predicción
features3.iloc[:,5:].head(5) # Codificación de la matriz de datos para el método de predicción
print('The shape of our features is:', features3.shape) # Verificación tamaño de la matriz codificada
#####
import numpy as np # Librería NUMPY para manipulación de matriz en números reales
labels3 = np.array(features3['Potencia']) # Columna de datos que vamos a predecir
features3= features3.drop('Potencia', axis = 1) # Eliminar la columna objetivo a predecir de la matriz general de datos fotovoltaicos
print('The shape of our features is:', features3.shape)
features3.to_csv('M_Entrenamiento_Codificada_noFO_mes3.csv')# Exportación matriz de datos sin la columna objetivo a predecir
features3 = np.array(features3) # Conversión de la matriz de datos a números reales para ingresar al método de predicción

The shape of our features is: (744, 31)
The shape of our features is: (744, 30)

```

Figura 2.19. Proceso de codificación de datos o variables mensuales

En la parte inferior de la figura se observa el número de columnas exportadas en el primer caso se tiene (744, 31) debido a que se importa la columna de potencia, pero luego se observa (744, 30) esto es debido a que se vuelve a exportar, pero sin la columna que se va a predecir. El pseudocódigo y la tabla de la importación de datos fotovoltaicos de prueba se muestran en el ANEXO 31 y 32.

### 2.5.10 Cálculo del Error Absoluto

Para saber si se cumplió el objetivo al que se quiere llegar el modelo se debe analizar el error existente de la potencia original con la resultante, este proceso se realizó al obtener el valor absoluto con la resta de la potencia predicha con la original. Y el cálculo de MAPE que es el error porcentual medio. La ecuación (2.7) es utilizada para el cálculo.

$$ABS\ ERROR = ABS (PREDICCIÓN - ORIGINAL) \quad (2.7)$$

Donde:

ABS ERROR: Es el error absoluto.

PREDICCIÓN: Es la potencia que el modelo predijo.

ORIGINAL: Es la potencia asumida del día con el que se entrena el modelo.

### **Error porcentual medio (MAPE)**

El MAPE es un porcentaje que me indica la exactitud del error, es más sencillo su interpretación que los demás métodos estadísticos que se aplican en medición de exactitud. Si el MAPE es de 4, el pronóstico se encuentra errado en el 4%. Los valores cercanos a 0 afectan el MAPE. El MAPE es la división entre el error absoluto y el dato real como indica la ecuación (2.8).

$$MAPE = 100 * \left( \frac{ABS\ ERROR}{DATO\ REAL} \right) \quad (2.8)$$

Donde:

MAPE: Error porcentual medio.

ABS ERROR: Es el error absoluto obtenido.

## Conclusiones del capítulo II

- Con el estudio en campo realizado se creó una base de datos de las variables de irradiación solar y temperatura del año 2020, estas variables fueron ordenadas tratadas de una manera adecuada para eliminar valores atípicos que son los que distorsionan el modelo de predicción.
- Con la constante búsqueda bibliográfica y la aplicación de la matriz de confusión se llegó a definir el modelo para utilizar en la predicción, llegando a elaborarlo con “árboles de decisión” por su facilidad al momento de crear su código y la solvencia con la que brinda buenos resultados.
- El modelo de predicción random forest o árboles de decisión toma solamente las variables que tienen mayor incidencia en la generación fotovoltaica, para que el modelo de predicción tenga mayor precisión y el tiempo de respuesta sea menor, también con el estado del arte que el algoritmo brinda se tiene una respuesta a sistemas con parámetros de comportamiento estocástico.

## CAPÍTULO III

### APLICACIÓN Y/O VALIDACIÓN DE LA PROPUESTA

Mediante el análisis teórico del capítulo I y la metodología plasmada en el capítulo II, se procede a realizar la validación del modelo, indicando en este capítulo los resultados obtenidos desde la elaboración del algoritmo dando a conocer que librerías se importó, como se entrenó el modelo hasta llegar a validar comparándolo con el dato real de predicción.

#### 3.1 Análisis de resultados

Mediante la toma de datos realizados con el piranómetro **Solar Power Meter SM206**, y comparadas con datos de la nasa y meteonorm se observó un pequeño margen de error, considerando que la base de datos de estos programas se encuentra hasta el 2 014 no están datos del 2 020 que es el año de estudio de la presente investigación. El mes con mayor irradiancia en el sector de Mulaló en la hacienda “La Campiña” es diciembre con (2,53 kWh/m<sup>2</sup>), mientras que en la mediación el que presenta una baja irradiancia es enero y junio con (2,11 kWh/m<sup>2</sup>), en la Tabla 3.1, se presenta la comparación entre esta base de datos medida con el piranómetro y los programas existentes en internet.

Tabla 3.1. Comparación con datos medidos

Año	Datos reales de un año		Nasa PVSyst 1983-2005		Meteonorm 2010-2014	
	Wh/m <sup>2</sup>	KWh/m <sup>2</sup>	Wh/m <sup>2</sup>	KWh/m <sup>2</sup>	Wh/m <sup>2</sup>	KWh/m <sup>2</sup>
Enero	211,22	2,11	163,7	1,637	170,6	1,706
Febrero	234,55	2,35	172,8	1,728	154,1	1,541
Marzo	216,25	2,16	194,7	1,947	165,2	1,652
Abril	234,54	2,35	171,6	1,716	157,6	1,576
Mayo	214,90	2,15	166,5	1,665	143,3	1,433
Junio	211,40	2,11	155,4	1,554	134,9	1,349
Julio	220,82	2,21	154,7	1,547	145	1,45
Agosto	244,66	2,45	157,8	1,578	152,4	1,524
Septiembre	236,15	2,36	156,3	1,563	151,6	1,516
Octubre	220,50	2,21	166,8	1,668	158,1	1,581
Noviembre	240,35	2,40	167,1	1,671	142,3	1,423
Diciembre	253,44	2,53	166,8	1,668	158,4	1,584



Para apreciar de mejor manera estos datos se elaboró una curva comparativa que se muestra en la Figura 3.1. Por la presente comparación e interpretación de datos se llega a la conclusión que el lugar es apto para implementar energía eléctrica a base de energía fotovoltaica considerando también que la potencia requerida es baja debido a que en la hacienda se desea encender ciertos electrodomésticos con el fin del confort y no perder su esencia rústica.

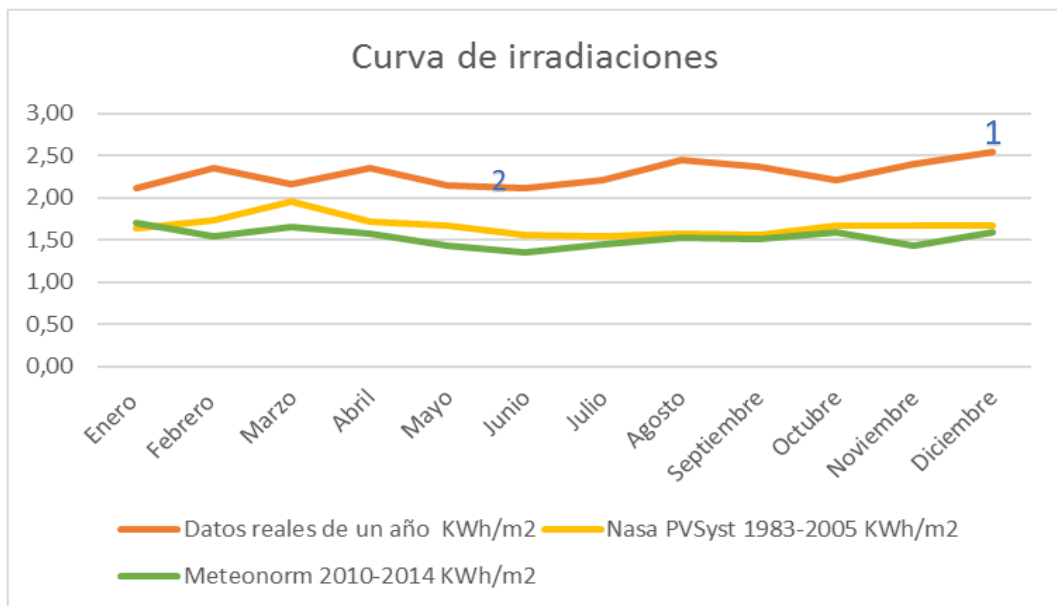


Figura 3.1. Curva de comparación de datos

En la relación de curvas de la Figura 3.1. entre la real tomada en campo y las existentes en base de datos de programas se llegó a analizar que el mes con mayor irradiación es diciembre que se encuentra en el punto 1, y el más bajo llegó a ser junio que es el punto 2, existiendo una diferencia de  $0,42 \text{ kWh/m}^2$ .

### 3.1.1 Análisis de las horas con mayor incidencia en el día

En el capítulo 2 se analizó el diagrama de frecuencias de irradiación totales, pero en el presente capítulo se interpreta las horas cuando más incidencia solar llega a la hacienda que va desde las 11h00 am hasta la 01h30 pm, donde la irradiación tiene un valor de  $127,5 \text{ W/m}^2$  considerando el mayor porcentaje de 32 %, y la más baja con un porcentaje del 10% como se muestra en la Figura 3.2. La Tabla para obtener esta curva se encuentra en el ANEXO 33, y mientras que el diagrama con las horas

con menor incidencia se encuentra en el ANEXO 34, y su respectiva tabla de análisis en el ANEXO 35.

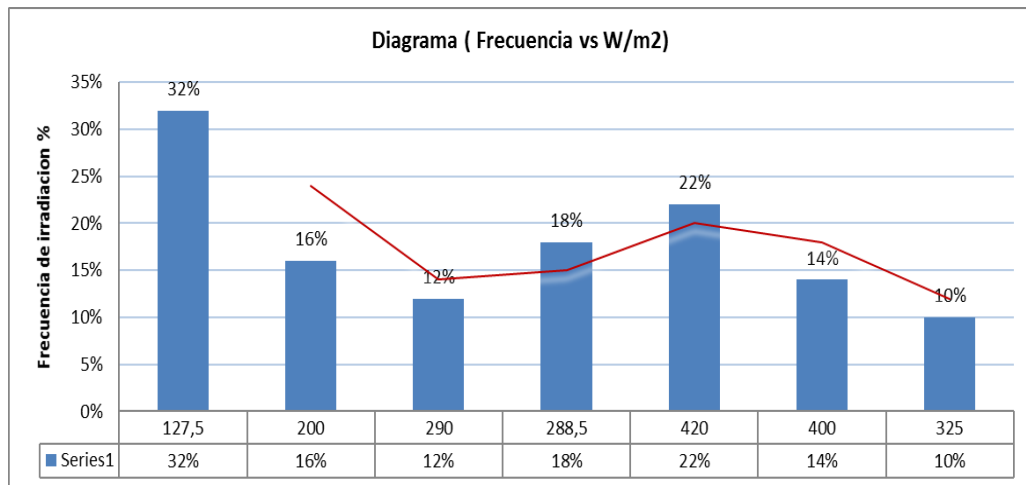

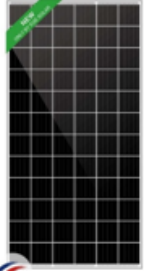


Figura 3.2. Diagrama (Frecuencia vs W/m2) de la mayor incidencia de irradiación solar

### 3.1.2 Toma de valores de potencia en la hacienda.

En la hacienda se realizó una toma de datos de potencia generada por un panel solar marca ECOGREEN ENERGY de 500 W en diferentes días y horas de cada mes para tener una base de datos utilizados en el modelo de predicción, estos datos importantes de acuerdo a la irradiación son: circuito abierto ( $V_{cc}$ ) y corriente de cortocircuito ( $I_{cc}$ ). El DataSheet del panel empleado se indica en la Tabla 3.2, mientras que los demás datos del panel se encuentran en ANEXO 36 y 37.

Tabla 3.2. DataSheet del panel fotovoltaico.

 	
Potencia de Salida ( $P_{max}$ )	530 W
Tolerancia de potencia	5 +/- W
Voltaje de circuito abierto $V_{oc}$	49 V
Corriente de Cortocircuito $I_{cc}$	13,76 A
Voltaje máximo ( $V_{mp}$ )	40,80 V
Corriente máxima ( $I_{mp}$ )	13.00 A
Eficiencia	20,70%

En la Tabla 3.3, se aprecia la potencia acumulada en un mes medida con un panel de 500 W marca ya nombrada en la Tabla 3.2.

Tabla 3.3. Potencia acumulada aproximada en un mes.

Hora	Potencia con panel de 500W (acumulada)
7:00:00	380,34
8:00:00	498,23
9:00:00	625,25
10:00:00	790,67
11:00:00	1025,39
12:00:00	1169,21
13:00:00	1235,23
14:00:00	1278,34
15:00:00	1278,23
16:00:00	702,87
17:00:00	297,34
18:00:00	207,3

Con la medición realizada se llega a conocer que la mayor generación que me brinda el panel empieza a la 13:00 pm con un valor máximo a las 14:00 pm de 1 278,34 esto debido a que esta es la hora pico donde más irradiación va a capta el panel por lo tanto va a brindarme una generación mayor que la medida en distintas horas, los datos se obtuvieron hasta las 18:00 debido a que el elemento de medición de esa hora en adelante ya no brindan valores. Los datos de la anterior tabla se interpretan con una mejor calidad en la Figura 3.3.

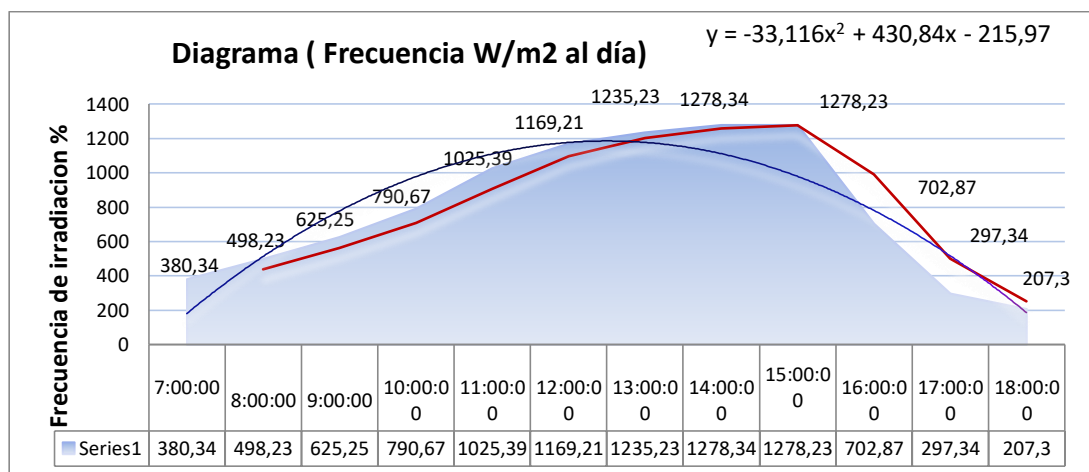


Figura 3.3. Diagrama (Frecuencia W/m2 al día)

Con la aplicación de las ecuaciones (2.4) y (2.5) planteadas en el capítulo II y los datos obtenidos en el sitio como se muestra en la Figura 3.4, se creó la Tabla 3.4, donde se analizó la potencia desde las 07:00 am hasta las 18:00 pm, con un promedio de datos por hora.

Tabla 3.4. Cálculo de la Energía asumiendo un panel fotovoltaico.

Hora	Irradiación (W/m <sup>2</sup> )	Potencia máxima generada	Eficiencia del Inversor	Energía (Wh)
7:00:00	380,34	102,5	0,9	92,2
8:00:00	498,23	140,1	0,9	126,1
9:00:00	625,25	180,5	0,9	162,5
10:00:00	790,67	233,3	0,9	209,9
11:00:00	1025,39	308,1	0,9	277,3
12:00:00	1169,21	353,9	0,9	318,5
13:00:00	1235,23	375,0	0,9	337,5
14:00:00	1278,34	388,7	0,9	349,8
15:00:00	1278,23	388,7	0,9	349,8
16:00:00	702,87	205,3	0,9	184,8
17:00:00	297,34	76,0	0,9	68,4
18:00:00	207,3	47,3	0,9	42,6
POTENCIA TOTAL GENERADA				2519,5



Figura 3.4. Datos obtenidos en la hacienda “La Campiña”.

Mediante el uso de la ecuación (2.4), se llegó a determinar que la máxima potencia en el día 20 de mayo se encuentra entre la hora 14:00 y 15:00 donde se ve una generación de irradiación de (1 278,23 W/m<sup>2</sup>) y por ende la potencia máxima es de (388,7 Wh). Con la ecuación (2.4), se llegó a determinar una generación de energía de (349,8 Wh) que llega a ser la máxima de este día.

### 3.1.3 Inicio y adecuación del programa

Este modelo fue creado en la plataforma Python en su plataforma de Google Colaboratory o “Colab”, esta plataforma ayuda a escribir y ejecutar código de programación de una manera adecuada, es de mucho ayuda para datos de aprendizaje, análisis de datos que es lo que se realizó. Al iniciar el programa lo primero que se realizó llamar a las librerías que fueron utilizadas en la elaboración del modelo. La Figura 3.5, indica las librerías que se cargaron.

```
import pandas as pd # Librería PANDAS para importar datos
from datetime import datetime
import matplotlib.pyplot as plt
import numpy as np # Librería NUMPY para manipulación de matriz en números reales
```

Figura 3.5. Librerías utilizadas.

### Lectura de datos

Los datos se importan de un documento creado en formato CSV que contiene los valores de las variables del año 2020, donde a la primera columna se le atribuye el año, la segunda el mes, la tercera el día, la cuarta es la hora esta columna lleva letras debido a que en la ejecución del programa van a ser llamados y con esta nomenclatura no me arrojó error en la importación, este proceso se plasmó debido al momento de la ejecución el programa iban presentando estos problemas, la cuarta es el tiempo en un intervalo de horas, y las columnas siguientes llevan las variables, como es el 01 de enero en el que observa todas las variables que son empleadas en el entrenamiento, como se muestra en la Tabla 3.5. Mientras que en la Tabla 3.6, es ejemplo de un resultado ejecutado con errores sin asignar un nombre a las horas. Los datos nulls deben ser tratados adecuadamente para no tener datos sin concordancia ni lógica.

Tabla 3.5. Datos que se cargaron al modelo.

	Year	Mes	Dia	Hora	Tiempo	Irradiacion	Temperatura	Potencia
0	2020	1	1	A	0	0.00	8.64	4.000000
1	2020	1	1	B	1	0.00	8.63	4.000000
2	2020	1	1	C	2	0.00	8.60	4.000000
3	2020	1	1	D	3	0.00	8.34	4.000000
4	2020	1	1	E	4	0.00	7.82	4.000000
5	2020	1	1	F	5	0.00	7.38	4.000000
6	2020	1	1	G	6	28.40	7.02	8.045584
7	2020	1	1	H	7	232.75	7.23	36.192254
8	2020	1	1	I	8	488.50	7.83	66.388250
9	2020	1	1	J	9	716.42	7.45	100.163758

Tabla 3.6. Datos erróneos de la ejecución del programa.

	Year	Mes	Dia	Tiempo	Irradiacion	Temperatura	Potencia
	24.0	24.0	24.0	24.000000	24.000000	24.000000	24.000000
	2020.0	1.0	1.0	11.500000	302.017083	9.627917	32.250962
	0.0	0.0	0.0	7.071068	391.672176	2.031433	36.225797
	2020.0	1.0	1.0	0.000000	0.000000	7.020000	4.000000
	2020.0	1.0	1.0	5.750000	0.000000	8.212500	4.000000
	2020.0	1.0	1.0	11.500000	16.300000	8.705000	6.221092
	2020.0	1.0	1.0	17.250000	639.680000	11.465000	65.210087
	2020.0	1.0	1.0	23.000000	1010.950000	13.360000	100.163758

Los datos faltantes es decir los nulls se fueron completando mediante la media aritmética teniendo en cuenta que en las horas de la noche no existe irradiación solar.

### **Variables importantes en la predicción**

En el modelo se emplearon tres variables las más importantes como: irradiación solar, temperatura y potencia, pero existieron otras variables utilizadas y mediante el pseudocódigo que se observa en la Figura 3.6, se dan un porcentaje y se tabula a

las variables mostradas en el capítulo II. En el ANEXO 38, se muestra una tabla con la totalidad de las variables.

```
# Import matplotlib for plotting and use magic command for Jupyter Notebooks
import matplotlib.pyplot as plt
%matplotlib inline
# Set the style
plt.style.use('fivethirtyeight')
x_values = list(range(len(importances))) # list of x locations for plotting
plt.bar(x_values, importances, orientation = 'vertical') # Make a bar chart
plt.xticks(x_values, feature_list1, rotation='vertical') # Tick labels for x axis
plt.ylabel('Porcentaje');
plt.xlabel('Variables');
plt.title('Importancia - Variables del Método de Predicción'),fontsize=18, color="BLACK", alpha = 1)
```

Figura 3.6. Pseudocódigo de variables.

### 3.1.4 Resultados de la ejecución del modelo de predicción diario

Después de entrenar el modelo con las variables planteadas en el capítulo II, se obtuvo una predicción similar a la real, para contextualizar el modelo se entrenó con las variables del día 01 y del día 20 pero al momento de ingresar las variables se elimina o no se le deja ver al programa la variable a predecir, para este proceso se debe entrenar adecuadamente los datos para que brinde un resultado cercano al real. En la Figura 3.7, se plantea el pseudocódigo que se ingresó para comparar las variables y en la Figura 3.8, se llega a graficar el resultado de la predicción en un gráfico de curvas realizado en el software empleado en la presente investigación.

```
# Gráfica Potencia Original VS Potencia Predecida
from datetime import datetime
import matplotlib.pyplot as plt
#plt.plot(fecha20,potencia_inicial20, 'r.-.', alpha = 0.4)
plt.plot(fecha20,potencia_inicial20,color="green", lw=0.5, ls='--', label = 'Potencia Original') #marker='o'
plt.plot(fecha20,predictions,color="blue", lw=0.5, ls='--', label = 'Potencia Predecida')
plt.legend()
plt.gcf().set_size_inches(10, 5)
plt.xticks(rotation = '60');
plt.xlabel('Fecha (Año/Mes/Día/Hora)',fontsize=16, color="black")
plt.ylabel('Potencia (W)', fontsize=16, color="black")
plt.title('POTENCIA DÍA 20'),fontsize=18, color="BLACK", alpha = 1)
plt.show()
```

Figura 3.7. Pseudocódigo de gráfico de potencia original vs potencia predicha.

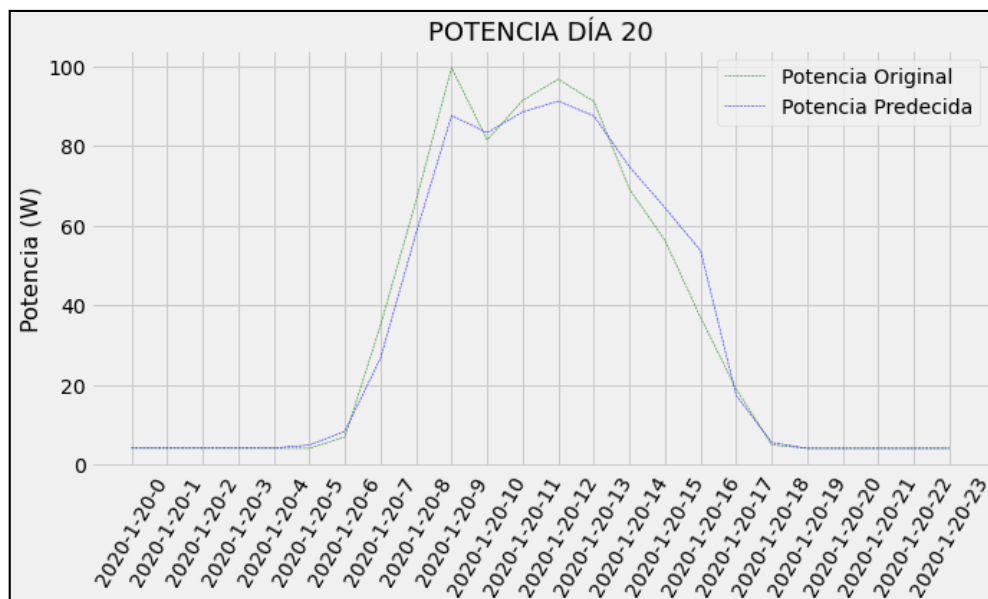


Figura 3.8. Gráfica de resultado.

La Figura 3.8. es el resultado obtenido con el modelo elaborado a base de machine learning y su librería random forest que tiene la capacidad de predecir generación fotovoltaica. La variable original se encuentra en la curva de color verde mientras que la variable de predicción se encuentra de color azul, como se aprecia el comportamiento de las curvas tiene una cierta diferencia una de la otra.

### Error absoluto

```
# Calculate the absolute errors
errors = abs(predictions - labels20)
print('Mean Absolute Error:', round(np.mean(errors), 2), 'Wattios.')
# Calculate mean absolute percentage error (MAPE)
mape = 100 * ((errors) / (labels20))
print(mape)
accuracy = 100 - np.mean(mape)
print('Accuracy:', round(accuracy, 2), '%.')
```

Figura 3.9. Gráfica de resultado.

La Figura 3.9, indica el proceso ingresado al software para determinar el MAPE o error absoluto, donde se va restando los datos de la variable de predicción con los de la variable original. Con la ejecución del pseudocódigo se llegó a obtener datos plasmados en la Tabla 3.7, donde se emplearon las fórmulas planteadas y detalladas en el capítulo II para encontrar el error. El error absoluto resultante es de 3,24 Wattios brindado un rendimiento del modelo de 91,41 %, con este porcentaje se concluye que el modelo está correctamente desarrollado y tiene un funcionamiento



adecuado. En la Figura 3.10, se corroboró los datos obtenidos en cálculo con los arrojados por el modelo.

Tabla 3.7. Tabla de cálculo del error.

HORA	P. ORIGINAL	P. PREDICCIÓN	ERROR ABS	MAPE
0:00	4	4,096164	0,096164	2,4041
1:00	4	4,096164	0,096164	2,4041
2:00	4	4,096164	0,096164	2,4041
3:00	4	4,096164	0,096164	2,4041
4:00	4	4,096164	0,096164	2,4041
5:00	4	4,096164	0,096164	2,4041
6:00	6,881766	8,312983	1,431217	20,7972343
7:00	34,889350	26,62386	8,265490	23,6905818
8:00	66,218391	58,04884	8,169551	12,3372841
9:00	99,668456	87,65973	12,008731	12,0486777
10:00	81,525365	83,337761	1,812396	2,22310688
11:00	91,496111	88,58982	2,906293	3,17641151
12:00	96,806394	91,27030	5,536098	5,71873176
13:00	91,156605	87,57222	3,584386	3,9321188
14:00	69,172156	74,78047	5,608313	8,10776087
15:00	56,259800	64,49330	8,233498	14,6347801
16:00	36,892668	53,679761	16,787093	45,5025183
17:00	18,945200	17,343239	1,601961	8,45576188
18:00	4,889518	5,495335	0,605817	12,390117
19:00	4,000000	4,000000	0,000000	0
20:00	4,000000	4,000000	0,000000	0
21:00	4,000000	4,000000	0,000000	0
22:00	4,000000	4,000000	0,000000	0
23:00	4,000000	4,000000	0,000000	0

```

Mean Absolute Error: 3.24 Watios.
[ 2.40409396 2.40409396 2.40409396 2.40409396 2.40409396 21.03302169
 20.79722728 23.69058202 12.33728419 12.04867819 2.22310614 3.17641102
 5.7187323 3.93211908 8.10776079 14.63478038 45.50251674 8.47630379
 12.3901128 0. 0. 0. 0. 0. ]
Accuracy: 91.41 %.
  
```

Figura 3.10. Error absoluto.

### 3.1.5 Resultados de la validación del modelo de predicción mensual

Mediante el análisis de los datos recolectados se planteó una base mensual similar a la base de datos diaria a diferencia que se aplicó la base de datos completa, en el proceso se identificó las relaciones puntuales entre las variables más utilizadas o las que mayor porcentaje de participación tienen para demostrar la eficiencia del modelo predictivo que se elaboró con random forest. El algoritmo para ser más preciso debe actuar sobre él diferentes variables en el análisis bibliográfico previo realizado, se describe que un modelo llegue a tener una efectividad del 99 % cuando

intervienen 5 variables, pero para ser ejecutado se puede entrenar y validar desde 3 variables, este modelo fue entrenado con 3 variables con un rendimiento del 92,95 %, en la Figura 3.11, se observa la curva planteada de la variación de los componentes principales, en el punto 1 se interpretó con tres variables un rendimiento del modelo ya aceptable y en el punto 2, con más variables llega a ser una predicción que ya entra en la exactitud de la real.

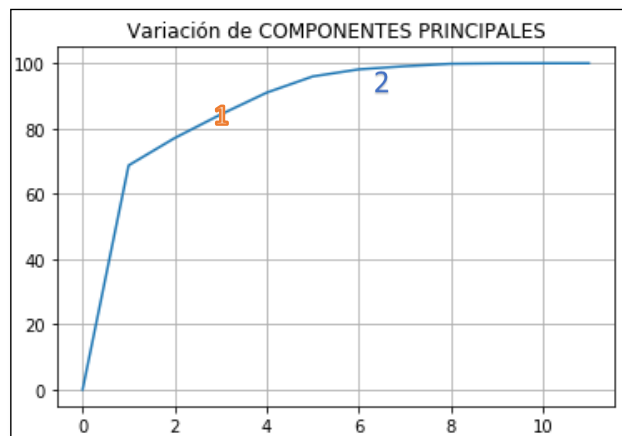


Figura 3.11. Variación de componentes principales.

La predicción mensual se realizó bajo el mismo procedimiento de la predicción diaria tomando los valores de variables de un mes cualquiera para entrenarle y del mes que se desea predecir la variable deseada se importa las demás variables y se le oculta la que se desea importar este proceso fue detallado en el capítulo II.

La predicción mensual elaborada es del mes 05 es decir mayo, se entrenó con los datos de marzo. Las curvas de irradiación solar y temperatura del mes de mayo se encuentran en ANEXO 39 y 40. El proceso de entrenamiento se elaboró con los mismos 1 000 árboles de decisión. En la Figura 3.12, se muestra el pseudocódigo para graficar la curva de la variable de predicción vs la variable real después de realizar los procesos como: codificación, entrenamiento.

```

# Gráfica Potencia Original VS Potencia Predecida
from datetime import datetime
import matplotlib.pyplot as plt
#plt.plot(fecha20,potencia_inicial20, 'r--', alpha = 0.4)
plt.plot(fecha5,potencia_inicial5,color="green", lw=0.7, ls='--', label = 'Potencia Original') #marker='o'
plt.plot(fecha5,predictions5,color="blue", lw=0.7, ls='--', label = 'Potencia Predecida')
plt.legend()
plt.gcf().set_size_inches(12, 6)
plt.xticks(rotation = '60');
#plt.xticks(range(0,700,100),fecha5)
plt.xticks(np.arange(0, 744, step=24))
plt.xlabel('Fecha (Año/Mes/Día/Hora)',fontsize=16, color="black")
plt.ylabel('Potencia (W)', fontsize=16, color="black")
plt.title('POTENCIA MES 5',fontsize=18, color="BLACK", alpha = 1)
plt.show()

```

Figura 3.12. Pseudocódigo de las curvas.

En la Figura 3.13, se plasma los resultados de la ejecución del pseudocódigo, arrojando dos curvas; la curva de color verde es la variable real mientras que la curva de color azul es la variable de predicción. Cabe recalcar que en la predicción mensual se exponen los 31 días correspondientes al mes de mayo, con esta predicción obtenida se puede realizar un análisis de para conocer si dicha irradiación va a cubrir la demanda solicitada.

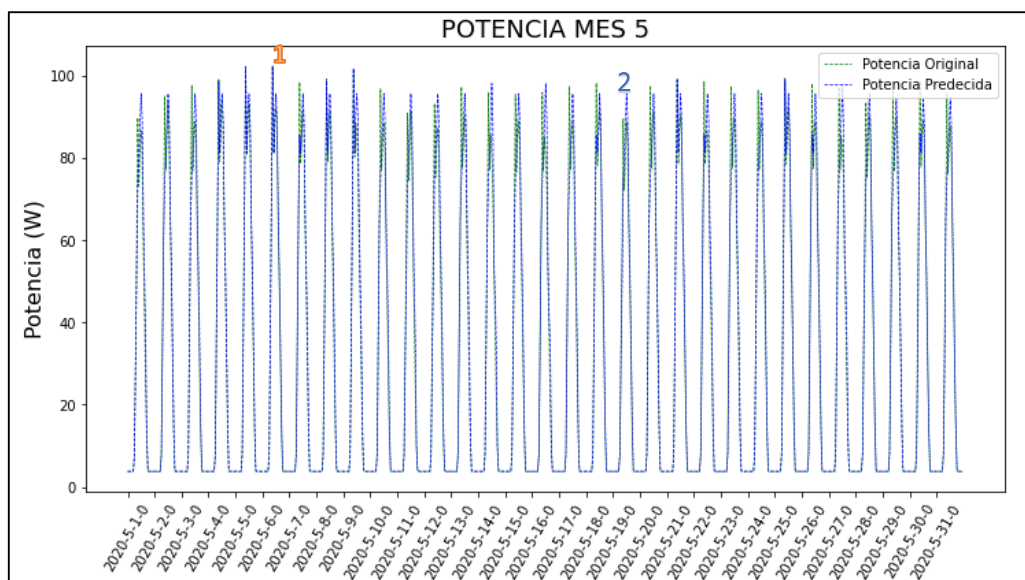


Figura 3.13. Potencia Original vs Potencia de Predicción.

1.- El punto más elevado en la predicción es el 06 de mayo arrojando una potencia de generación sobre los 97 W.

2.- El punto más bajo en la predicción aparece en el día 19 de mayo con una potencia de generación bajo los 95W.

El modelo de predicción basado en Árboles de decisión creó una curva por cada día del mes, empieza en el 01 hasta el último día del mes de mayo que es el 31. Mientras se ejecutó el modelo se fue probando varias predicciones como se muestran en los ANEXOS 41, 42 Y 43, hasta que el error resultante esté dentro de los parámetros adecuados y permisibles.

### Error de predicción mensual

```
# Calculate the absolute errors
errors5 = abs(predictions5 - labels5)
print('Mean Absolute Error:', round(np.mean(errors5), 2), 'Wattios.')
# Calculate mean absolute percentage error (MAPE)
mape5 = 100 * ((errors5) / (labels5))
print(mape5)
accuracy5 = 100 - np.mean(mape5)
print('Accuracy:', round(accuracy5, 2), '%')
```

Figura 3.14. Error de predicción mensual.

La Figura 3.14, muestra el pseudocódigo para obtener el error absoluto y el MAPE del modelo mensual, donde se llegó a una precisión del 94,47% de error, es decir la potencia de prueba colocada tuvo un margen de error de 2,37 Wattios, por ende el rendimiento mensual se incrementó a comparación diario al 94,47%.

Mean Absolute Error: 2.37 Wattios.

Accuracy: 94.47 %.

Por la magnitud de datos de predicción la tabla se plantea en ANEXO 44.

### 3.1.6 Validación cruzada del modelo de predicción (Cross-Validation)

Para estimar el desempeño del modelo predictivo se utilizó la Validación Cruzada, para la predicción se dividió el proceso en dos partes entrenamiento e implementación. La primera parte es exclusiva para entrenar el modelo y en la implementación se aplicó datos que no son observados por el algoritmo para

conocer el resultado en la implementación real, para este algoritmo se utilizó 1 000 árboles de decisión como se explica en la investigación.

En el proceso de creación se fueron aplicando datos de prueba para llegar a un desempeño de investigación idóneo. Primero se creó un modelo con 100 Árboles de decisión, pero el desempeño de entrenamiento fue de 0,8012 y el desempeño de implementación fue más bajo aún con un 0,7354, motivo por el cual el error es muy amplio y no se llegó a una predicción real del modelo. La validación cruzada consta en dividir la tabla de las variables de ingreso en partes iguales, en el caso de la investigación se llegó a dividirse en 1 000. En la Tabla 3.8 se muestran los datos obtenidos de la evaluación. Donde el mínimo valor 1 se considera 100 y el máximo son los 1 000 árboles utilizados.

Tabla 3.8. Tabla de evaluación de validación cruzada.

VALIDACIÓN CRUZADA		
EVALUACIÓN	K	RESULTADO
EVA	10	0,50
EVA	9	0,51
EVA	8	0,50
EVA	7	0,49
EVA	6	0,52
EVA	5	0,50
EVA	4	0,51
EVA	3	0,50
EVA	2	0,49
EVA	1	0,52
PROMEDIO		0,504

Para la validación del modelo de predicción se utilizó un valor seteado de 1 000 y con los datos de las variables de cada mes aplicando el proceso mencionado.

K: se debe aplicar valores de 5 o 10 que son los más recomendados debido a que si aumenta K aumenta proceso de entrenamiento y por ende tiene al aumento el desempeño del modelo. Este proceso se utilizó para la predicción por día debido a que los datos no son en gran magnitud.

El modelo final fue entrenado con toda la tabla de datos, el número 1 000 indica las rondas que se entrenó el modelo, este proceso se realizó para ver el número de rondas totales que se va a utilizar en el modelo. En la Figura 3.15, se encuentra la curva de entrenamiento, esta incrementa debido a la magnitud de datos. Y también se observa la curva de desempeño de predicción esta curva se va a estabilizando debido a que se llegó al número de rondas necesarias para el modelo de predicción.



Figura 3.15. Curvas de desempeño.

### **Conclusiones del capítulo III**

- Mediante los datos obtenidos de irradiación solar en el lugar de la investigación se determinó el mes que van a tener el pico más elevado como es agosto y el más bajo que es julio.
- La presente investigación tiene como resultados dos simulaciones la primera es la predicción diaria del día 20 de enero siendo entrenado con el día 01 del mismo mes, mientras que la segunda predicción mensual se entrenó con el mes de marzo para dar como resultado una predicción de mayo.
- Con la validación cruzada se llegó a conocer la exactitud de árboles de decisión con los que se elaboró el modelo, mediante varias pruebas erróneas se llegó a determinar 1 000 árboles ya que me permitió ingresar la totalidad de la base de datos obtenida.

## CONCLUSIONES

- Tras el análisis bibliográfico se llegó a seleccionar el algoritmo para la predicción como lo es árboles de decisión, debido a que textualmente analizado trabaja con datos estocásticos y su interpretación es simple. Y mediante el análisis comparativo con dos algoritmos predictivos con la herramienta matriz de confusión es el más cercano a la variable real con las siguientes medidas de rendimiento: precisión de 0,4780, exactitud de 0,4791 y la función F1 de 0,4486.
- Como se vino observando en el desarrollo de la investigación la irradiación llega de forma variable donde los meses con mayor incidencia son marzo 623,70 ( $W/m^2$ ) y septiembre con 629,53 ( $W/m^2$ ). Se consideró una irradiación de 7 horas captadas para llegar a obtener un factor de planta del 35 % y en el mejor escenario se consideró 12 horas de incidencia para tener un factor de planta del 50,4 %.
- Se llegó a solventar el objetivo específico la creación del algoritmo con el cual se elaboró dos predicciones una diaria y una mensual, siendo la más acertada la mensual que tiene un rendimiento del 94,4 % a comparación de la predicción diaria que tiene un rendimiento del 91,41 %, esto es debido a que en la predicción mensual se ingresa al proceso de entrenamiento la base de datos completa y el resultado depende de un entrenamiento pertinente del 80 %.
- El proceso que se cumplió para llegar a una predicción diaria consta en ingresar las variables de irradiación, temperatura y potencia de prueba del 01 de enero y predecir la potencia de prueba del día 20 del mismo mes en la comparación con la variable real se tiene un error mínimo de 3,24 watos. En la predicción mensual se ingresó en el entrenamiento las variables de irradiación, temperatura y potencia de prueba del mes de marzo para predecir la potencia de prueba del mes de mayo dándome un error mínimo de 2,37 watos comparado con la variable real, llegando a concluir que el



algoritmo brinda una predicción la cual solventa los objetivos planteados en esta investigación de una manera satisfactoria.

## RECOMENDACIONES

- Dentro de la investigación muy exitosa como lo fue la elaborada, se desea extender los estudios expuestos con algoritmos de predicción para llegar a crear un modelo de optimización completo que en su predicción tenga un rendimiento elevado y el resultado sea similar a la variable real.
- Tomando en cuenta la implementación que se está dando de las energías renovables, se recomienda obtener datos eólicos para realizar una predicción con el algoritmo creado, para cuando se dé una implementación de un modelo solar y no cubra la carga por factores externos al diseño se complemente con la energía eólica.
- En base a los resultados obtenidos y los escenarios que se plasmaron con la recopilación de datos, se recomienda alargar la investigación siendo el principal tema de estudio las consecuencias y el problema que traería la emisión de ceniza por parte del volcán Cotopaxi en sistemas fotovoltaicos, en qué porcentaje disminuiría la irradiación solar en el ambiente y que soluciones se podría dar para contrarrestar este caso fortuito.
- Con la elaboración del algoritmo es recomendable en estudios posteriores la implementación de nuevas variables que pueden ayudar a aumentar el porcentaje de rendimiento del modelo y explicar el comportamiento de la irradiación solar, valerse de variables adicionales como velocidad del viento, humedad relativa, punto de rocío entre otros que puede ser de gran aporte para robustecer el modelo.

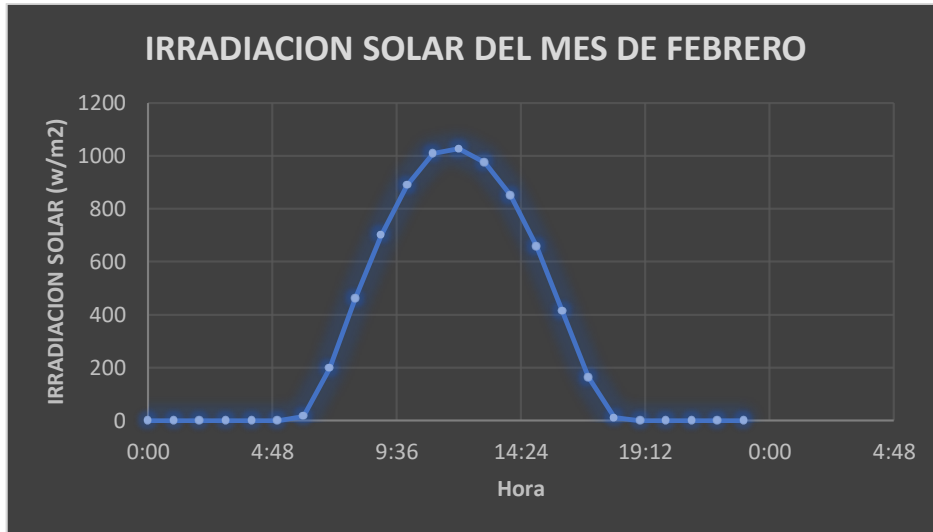
## REFERENCIAS BIBLIOGRÁFICA

- [1] D. Guasch, “Modelado y Análisis de Sistemas Fotovoltaicos,” p. 224, 2006.
- [2] J. Crespo and J. Peralta, “Estudio e implementación de un sistema de micro generación solar fotovoltaico para autoconsumo (caso de estudio considerando la regulación Nro. ARCONEL- 003/18),” 2019.
- [3] M. Ruiz and M. Garcia, “Interoperabilidad entre Medidores Inteligentes de Energía Eléctrica Residencial,” *Prim. Congr. Int. y Expo Científica*, no. January 2014, p. 101, 2015.
- [4] R. S. Carlos and D. H. Milone, “Optimización de centrales fotovoltaicas mediante algoritmos evolutivos,” pp. 252–263, 2012.
- [5] A. R. Criollo Ríos, “Determinación del tamaño óptimo de los sistemas de almacenamiento de energía BESS en sistemas fotovoltaicos conectados a la red eléctrica, mediante programación lineal entera mixta,” p. 122, 2017.
- [6] T. Y. Frías, “Colector Térmico Fotovoltaico Interconectados a la Red,” 2017.
- [7] M. Beltrán, Aurelio Morera Hernández, F. E. López, and R. Villela, “Prospectiva De Las Energías Eolica Y Solar Fotovoltaica,” *Cienc. UAT*, vol. 11, no. 2, pp. 105–117, 2017.
- [8] J. I. López Sánchez and L. E. Carretero Díaz, “La inteligencia artificial y la ingeniería del conocimiento como soporte para las técnicas de decisión basadas en la gestión del conocimiento,” *Dir. y Organ. Rev. Dir. Organ. y Adm. Empres.*, pp. 171–185, 2018.
- [9] A. Kumar and S. P. Panda, “A Survey: How Python Pitches in IT-World,” *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019*, pp. 248–251, 2019, doi: 10.1109/COMITCon.2019.8862251.
- [10] M. A. Abella, F. C. Romero, and S. D. B. Fotovoltaico, “Máster en Energías Renovables y Mercado Energético Energía Solar Fotovoltaica Escuela de Organización Industrial Sistemas de Bombeo Fotovoltaico,” *Dimens. Sist. Fotovoltaicos*, vol. 1, no. 1, p. 73, 2017.
- [11] Héctor L. Gasquet, “Conversión de la Luz Solar en Energía Eléctrica,” *Solartronic*, vol. 52, no. 90, p. 196, 20017.
- [12] M. Serrano, “Dimensionado de sistemas fotovoltaicos aislados,” *Univ. Jaén*, no. January, p. 338, 2018.
- [13] C. A. La, R. E. D. Eléctrica, E. N. La, C. Omar, and B. Avila, *Buele, C. Chazi, A.* 2019.
- [14] H. Ramos and R. Luna, “Diseño de un sistema fotovoltaico integrado a la red para el área de estacionamiento de la Universidad Tecnológica De Salamanca,” 2016.

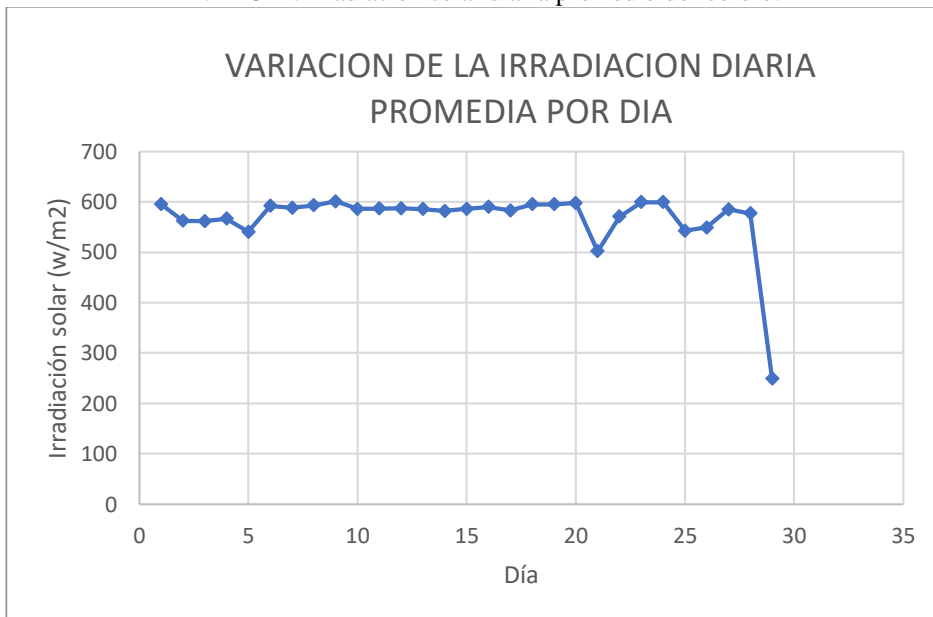
- [15] L. O. Perpiñan, “Energía S Olar,” no. Mayo, 2014.
- [16] CONELEC, “Atlas solar del ecuador,” *Conelec*, pp. 1–51, 2008.
- [17] D. Vaca, “Validación de datos satelitales de radiación solar utilizando mediciones terrestres para el Ecuador,” p. 125, 2018.
- [18] M. S. Alasa and E. P. Araujo, “Grado en Ingeniería en Tecnologías Industriales Predicción de producción de energía solar a partir de inteligencia artificial Escuela Técnica Superior de Ingeniería Industrial de Barcelona,” 2020.
- [19] J. M. Romero Rodríguez, “Modelo para predicción de potencia de paneles fotovoltaicos utilizando técnicas de clasificación no supervisada y redes neuronales artificiales,” 2020.
- [20] C. Ferrer Martínez, J. Chilo, and V. A. Terré, “UNIVERSITAT POLITÈCNICA DE VALÈNCIA ‘Machine Learning for Solar Energy Prediction’ TRABAJO FINAL DE GRADO,” 2018.
- [21] R. Barrientos, “Árboles de decisión como herramienta en el diagnóstico Médico,” *Artic. Orig.*, pp. 20–24, 20016.
- [22] J. Trujillano, J. March, and A. Sorribas, “[Methodological approach to the use of artificial neural networks for predicting results in medicine].,” *Med. Clin. (Barc.)*, vol. 122 Suppl, pp. 59–67, 2004, doi: 10.1157/13057536.
- [23] J. Peña, “SOLAR A PARTIR DE REDES NEURONALES ANDRÉS FELIPE GONZÁLEZ RODRÍGUEZ Trabajo de grado para optar al título de Ingeniero mecatrónico JUAN ALEJANDRO PEÑA PALACIO ( PhD ),” 2013.
- [24] M. O. Sosa and E. C. Bruchmann, “Aportes esperados de la técnica de árboles de desición al aplicarlos a datos generados con la metodología blended learning,” p. 6, 2015.
- [25] C. alexis Pacheco cueva and roberto C. Herrera Albarracin, “Universidad técnica de cotopaxi,” *Univ. técnica cotopaxi*, vol. 1, p. 101, 2021.
- [26] C. P. Fabara, “Prediction of Generation in a Photovoltaic System through the application of Data Mining techniques Predicción de la Generación para un Sistema Fotovoltaico mediante la aplicación de técnicas de Minería de Datos,” no. 16, pp. 69–78, 2019.
- [27] Esteban Albán Silva – David Alemán Quimbiulco, “Prediccion De La Demanda Electrica Utilizando Modelos De Aprendizaje,” p. 209, 2016.

# ANEXOS

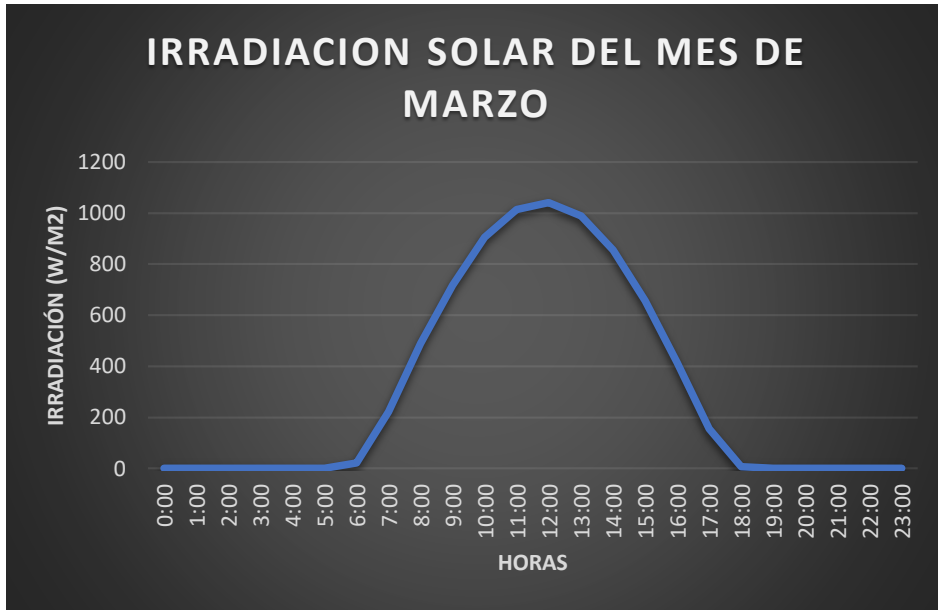
ANEXO 1: Curva de irradiación solar de febrero.



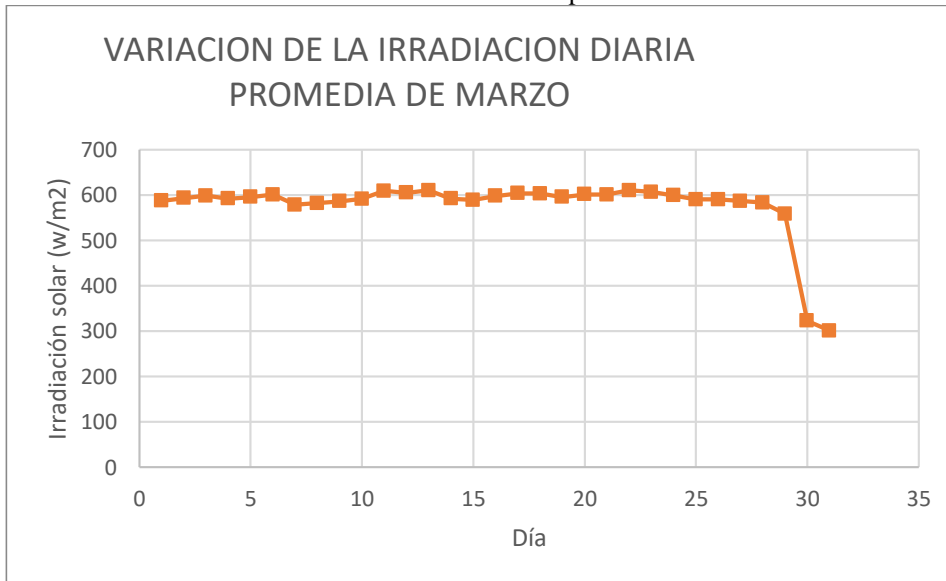
ANEXO 2: Irradiación solar diaria promedio de febrero.



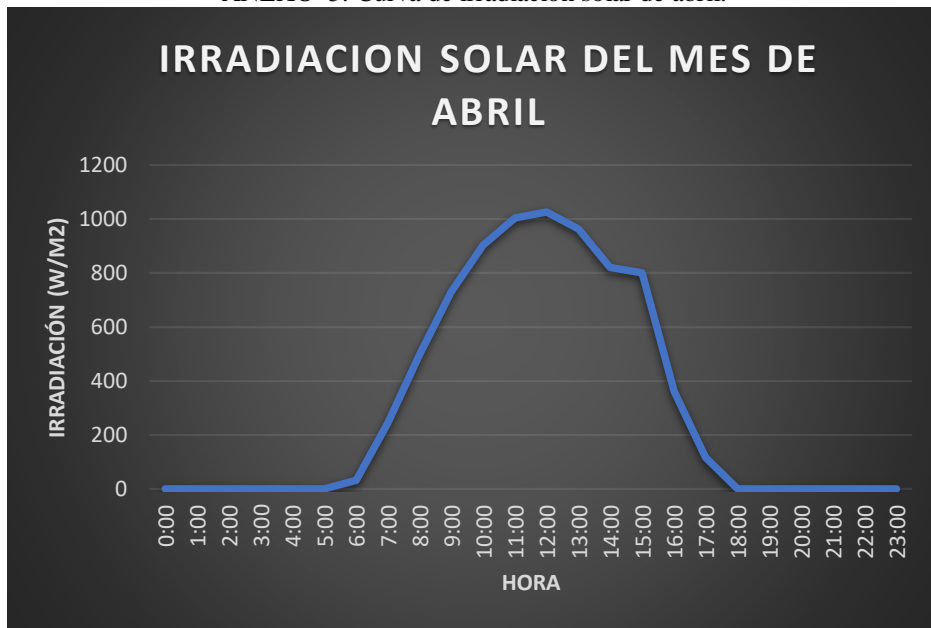
ANEXO 3: Curva de irradiación solar de marzo.



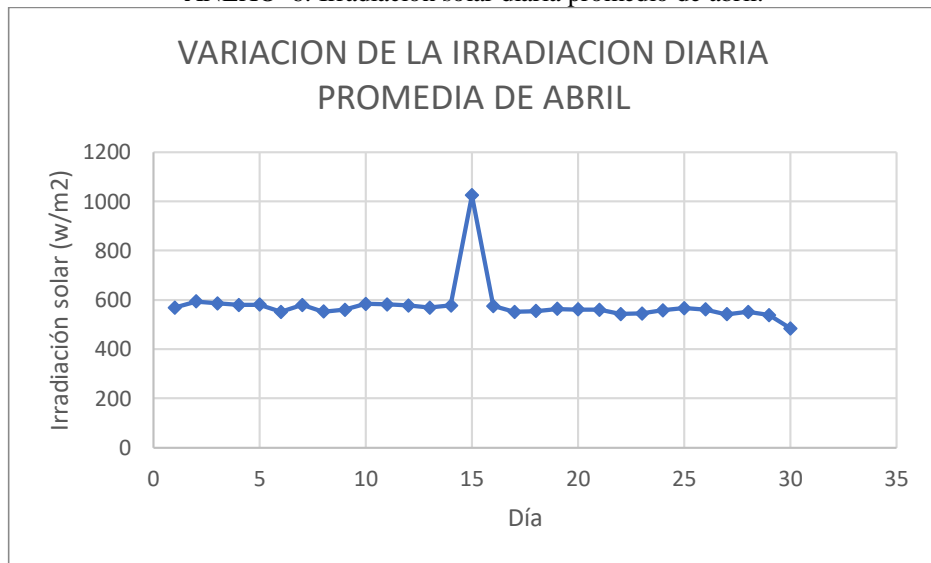
ANEXO 4: Irradiación solar diaria promedio de marzo.



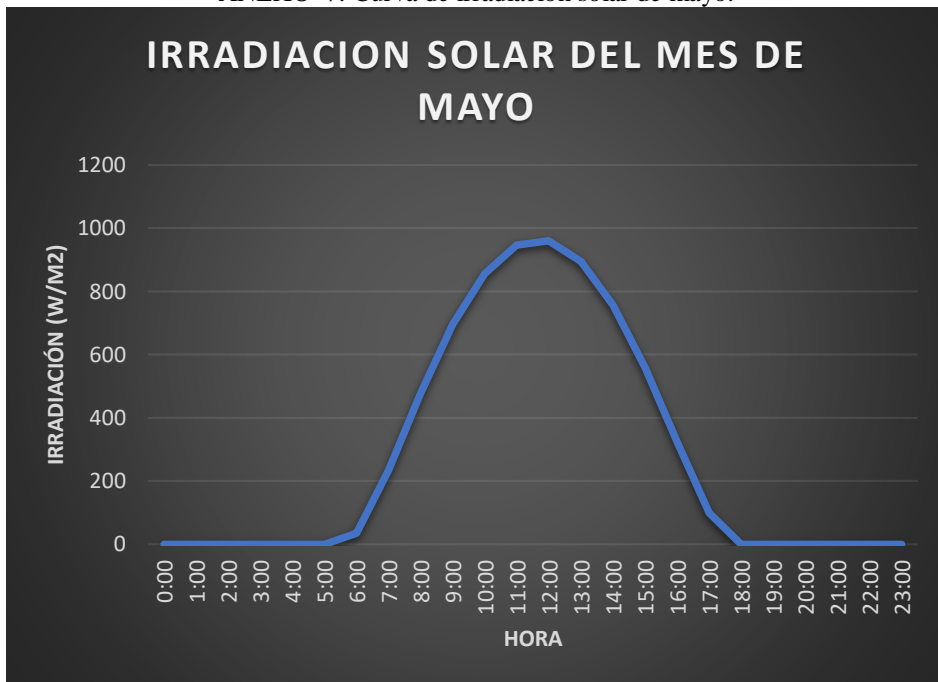
ANEXO 5: Curva de irradiación solar de abril.



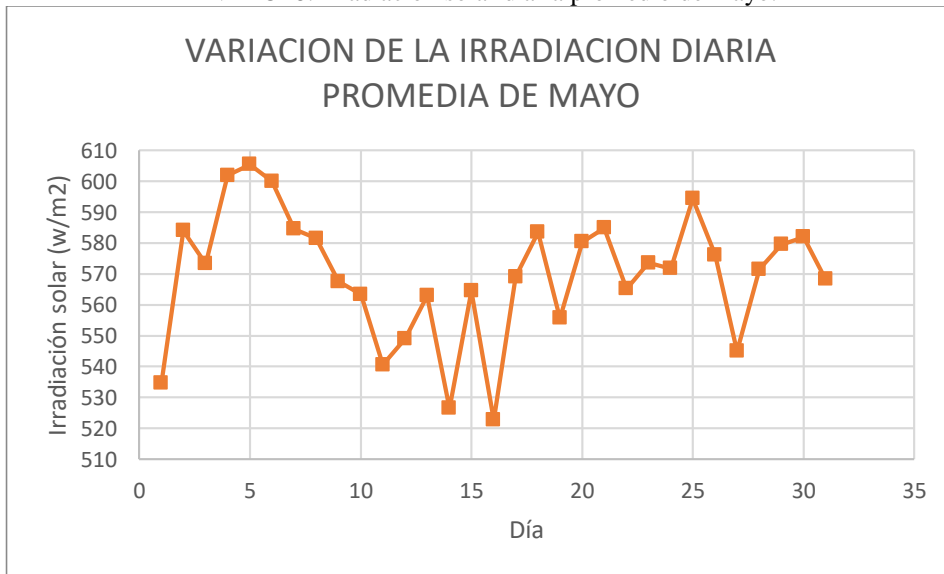
ANEXO 6: Irradiación solar diaria promedio de abril.



ANEXO 7: Curva de irradiación solar de mayo.

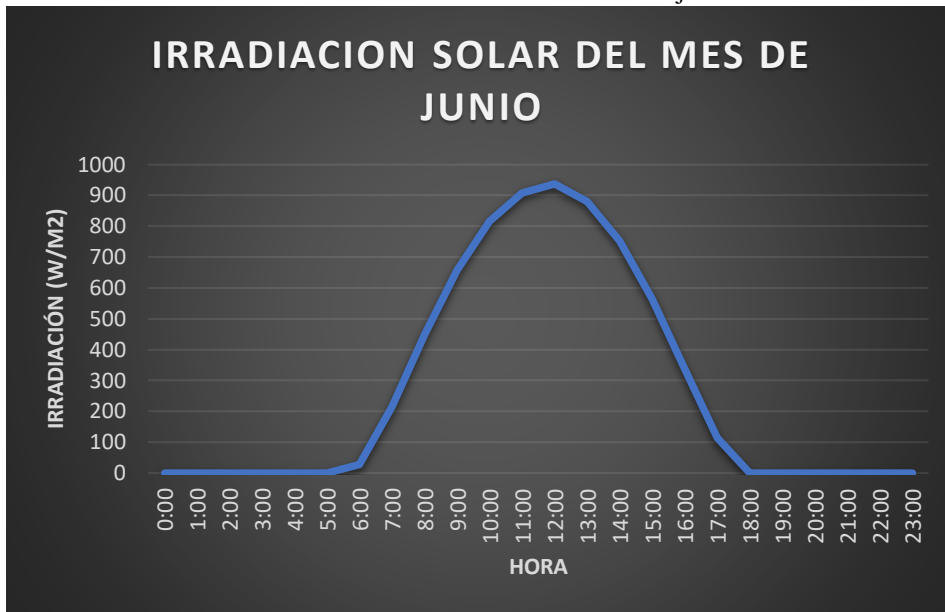


ANEXO 8: Irradiación solar diaria promedio de mayo.

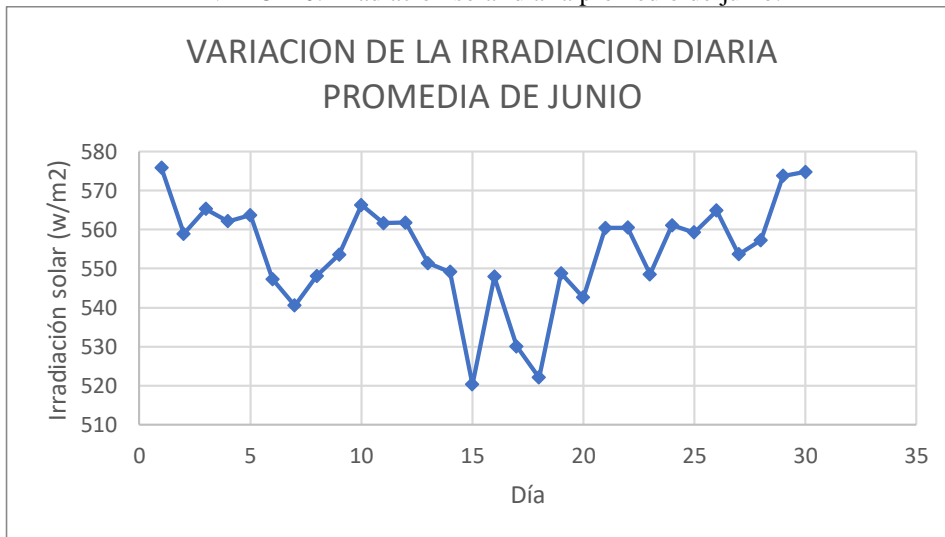




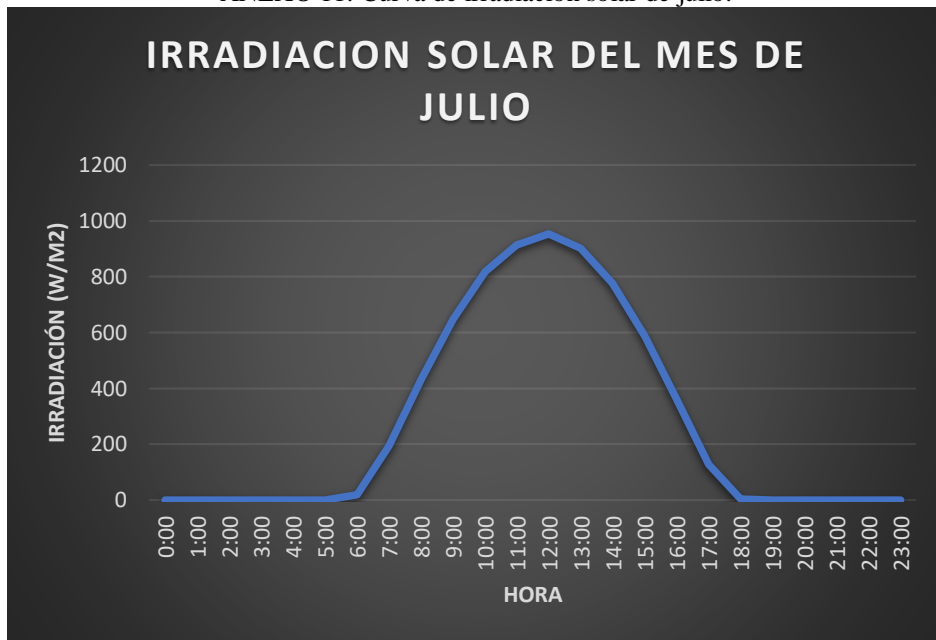
ANEXO 9: Curva de irradiación solar de junio.



ANEXO 10: Irradiación solar diaria promedio de junio.



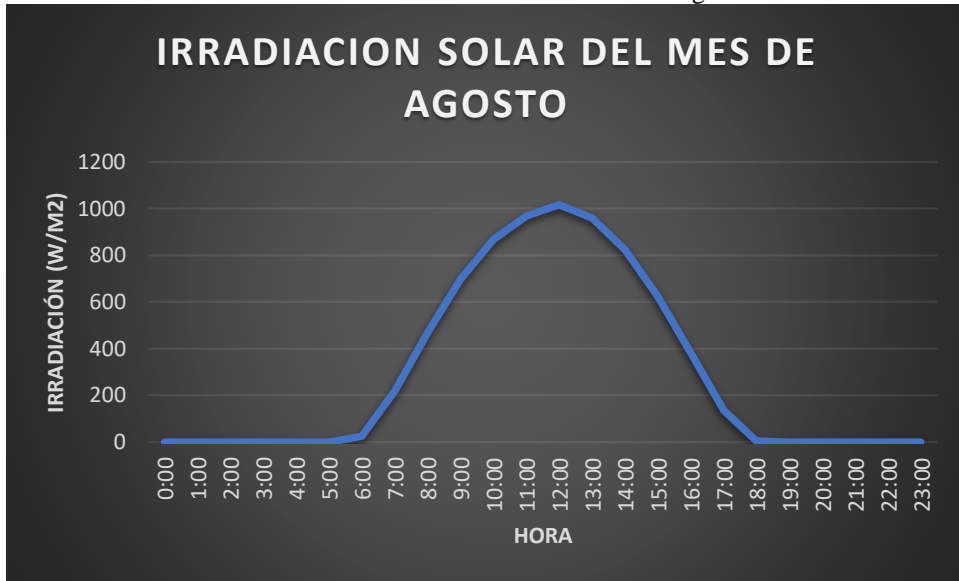
ANEXO 11: Curva de irradiación solar de julio.



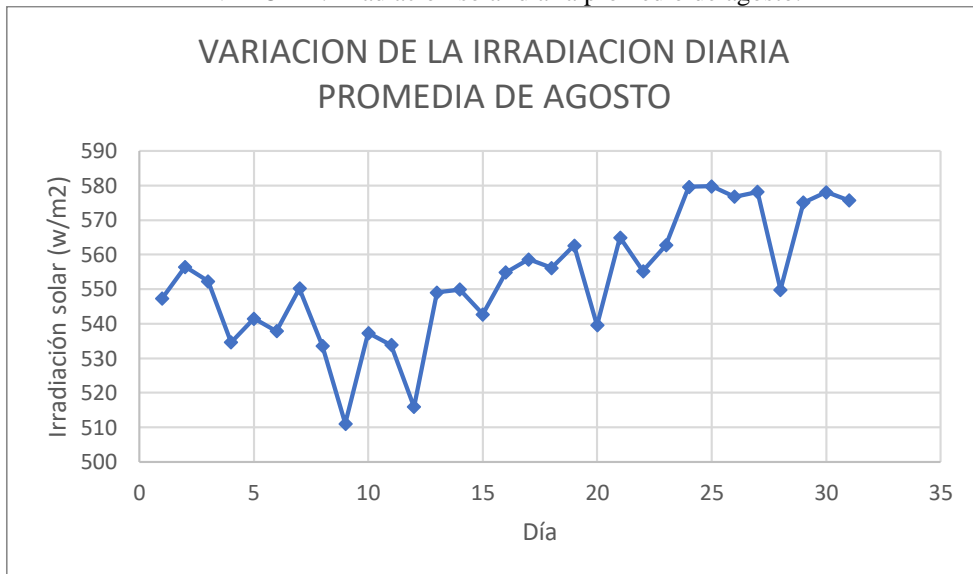
ANEXO 12: Irradiación solar diaria promedio de julio.



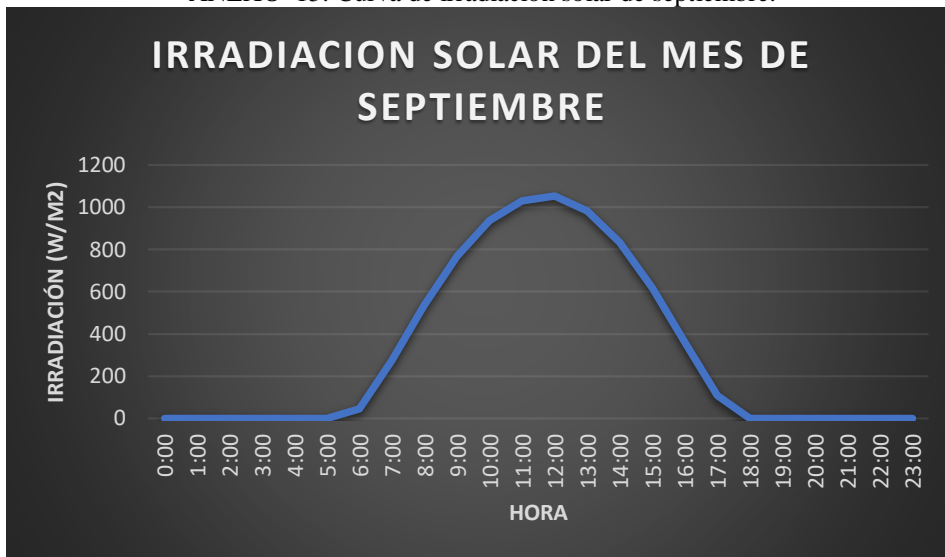
ANEXO 13: Curva de irradiación solar de agosto.



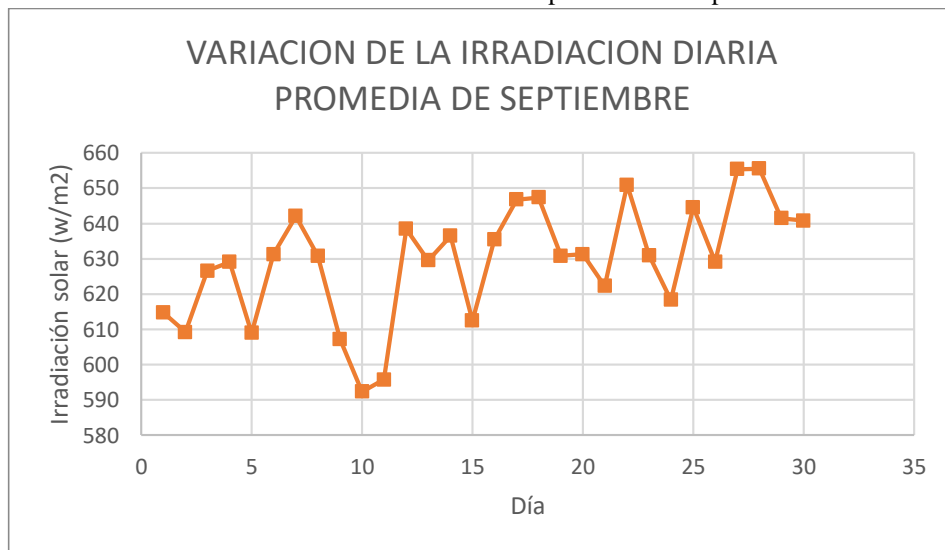
ANEXO 14: Irradiación solar diaria promedio de agosto.



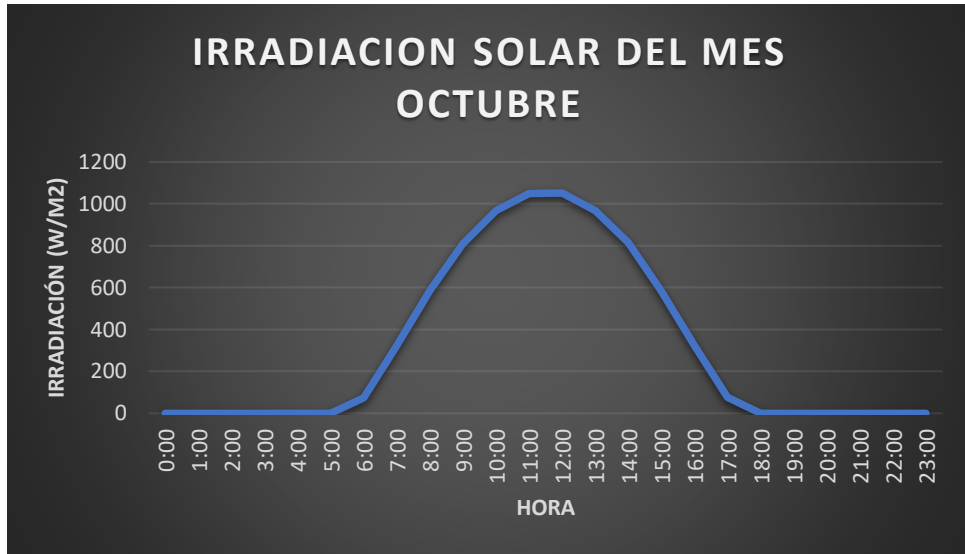
ANEXO 15: Curva de irradiación solar de septiembre.



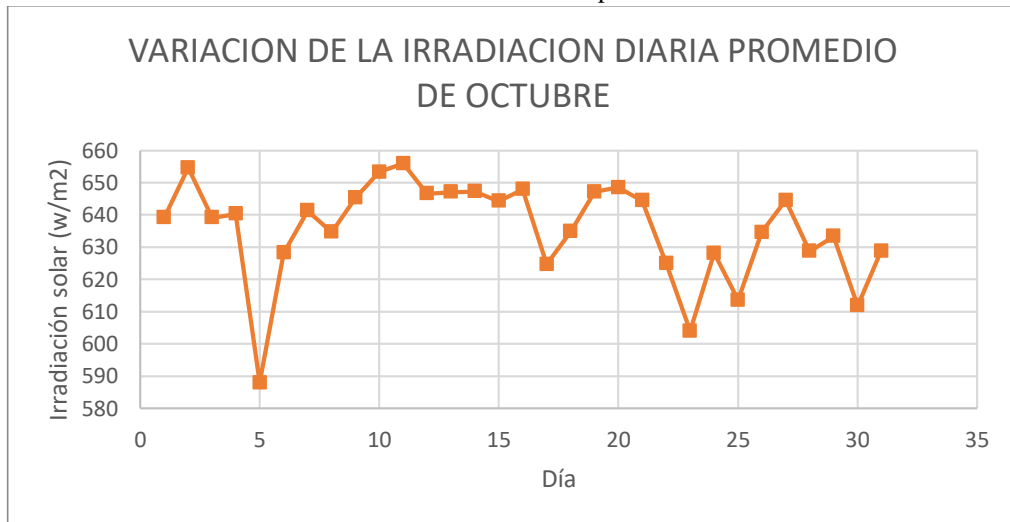
ANEXO 16: Irradiación solar diaria promedio de septiembre.



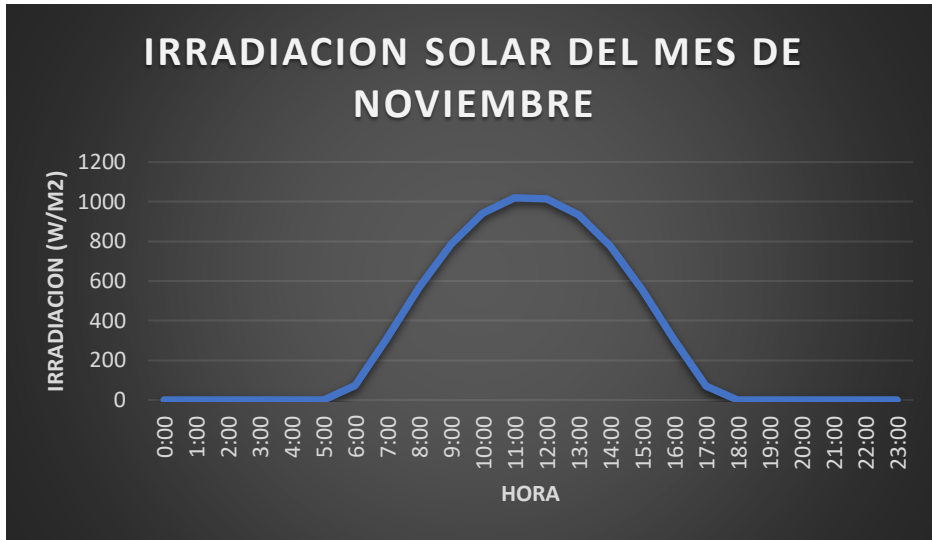
ANEXO 17: Curva de irradiación solar de octubre.



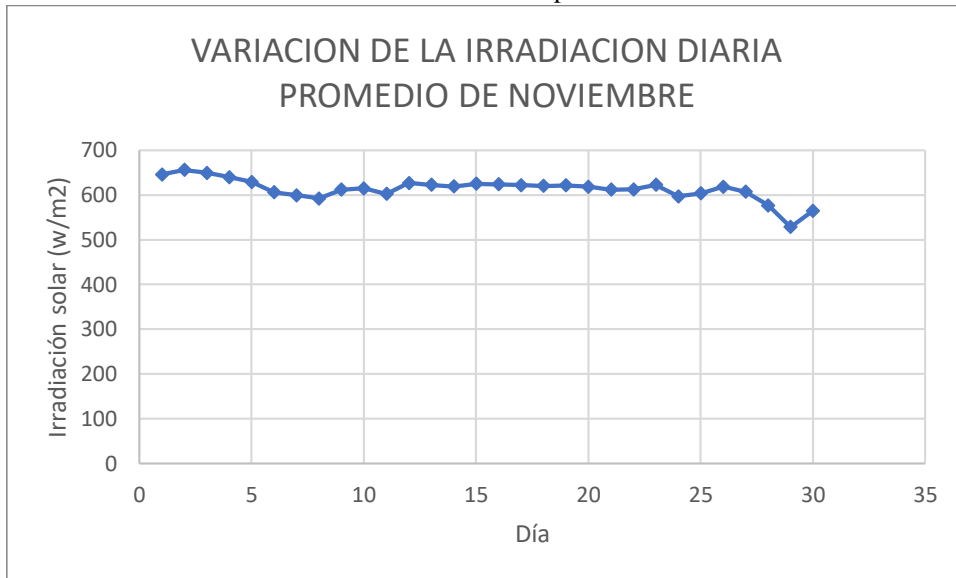
ANEXO 18: Irradiación solar diaria promedio de octubre.



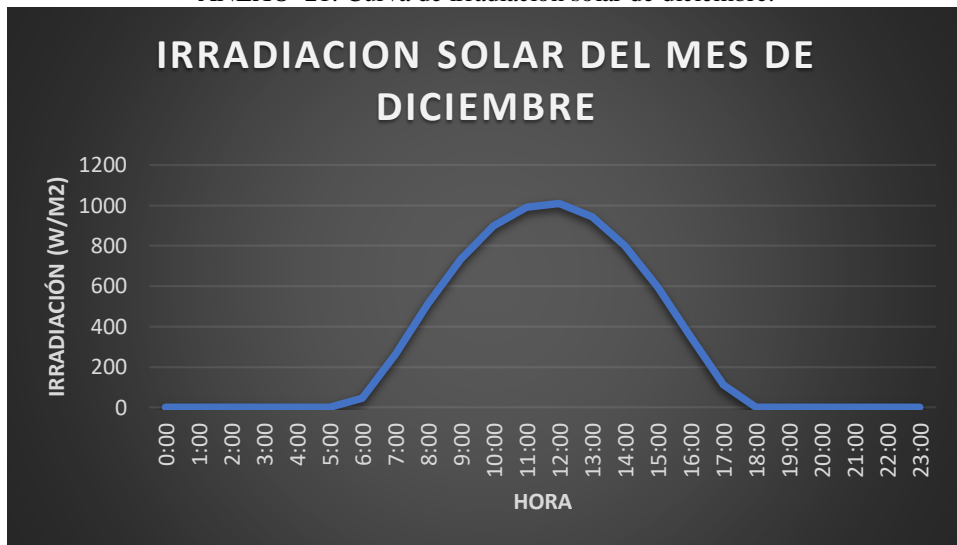
ANEXO 19: Curva de irradiación solar de noviembre.



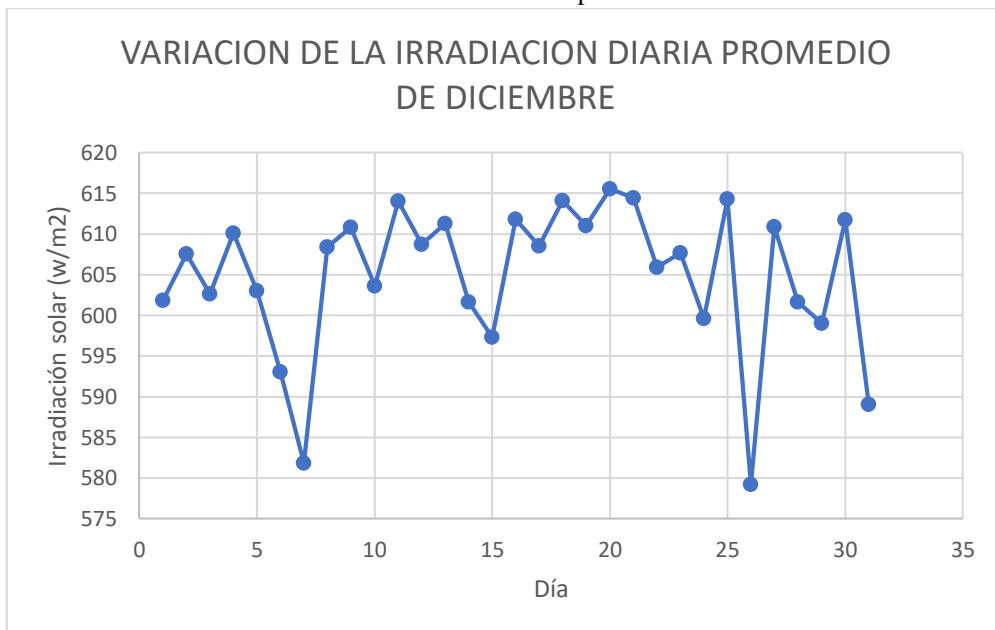
ANEXO 20: Irradiación solar diaria promedio de noviembre.



ANEXO 21: Curva de irradiación solar de diciembre.



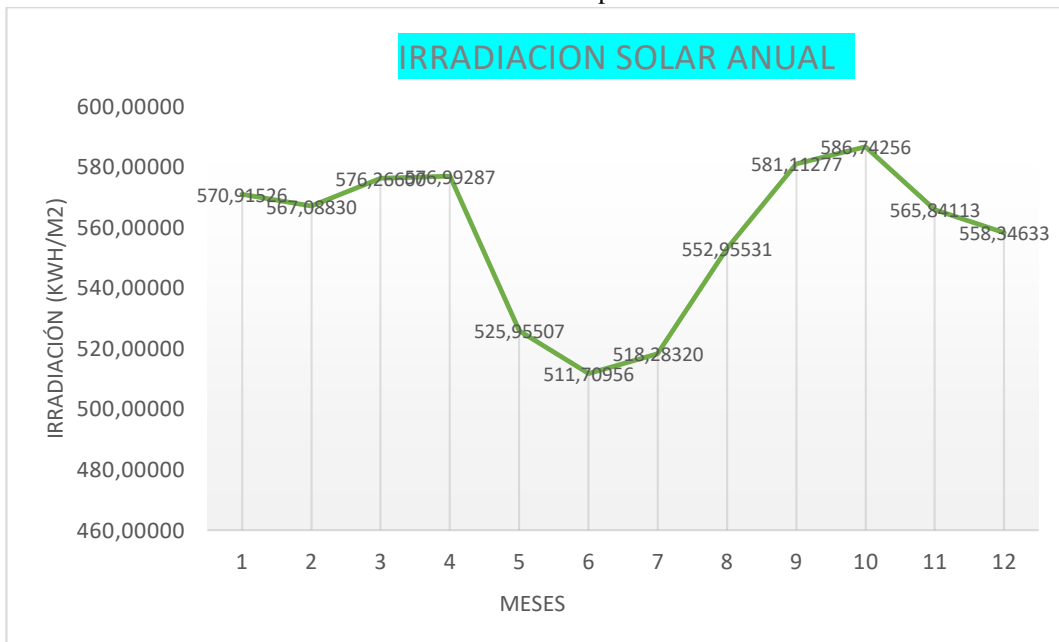
ANEXO 22: Irradiación solar diaria promedio de diciembre.



ANEXO 23: Irradiación promedio de cada mes.

IRRADIACION SOLAR PROMEDIO DEL AÑO POR HORAS												
	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOSTO	SEPTIEMBRE	OCTUBRE	NOVIEMBRE	DICIEMBRE
0:00	0	0	0	0	0	0	0	0	0	0	0	0
1:00	0	0	0	0	0	0	0	0	0	0	0	0
2:00	0	0	0	0	0	0	0	0	0	0	0	0
3:00	0	0	0	0	0	0	0	0	0	0	0	0
4:00	0	0	0	0	0	0	0	0	0	0	0	0
5:00	0	0	0	0	0	0	0	0	0	0	0	0
6:00	21,76276	15,6851724	22,0503226	32,1196667	35,2448387	27,2306667	19,8767742	24,6906452	45,25	74,3870968	73,68	44,3522581
7:00	212,04966	197,608276	221,786774	243,605333	234,816129	213,563333	194,326452	218,316452	273,715333	322,19871	311,285667	261,397742
8:00	467,48862	461,539655	489,467097	497,924333	477,094839	446,920333	431,237742	469,774839	538,897	586,550645	566,988	514,020968
9:00	697,52828	701,094483	718,096129	729,787333	694,393548	657,250667	648,590645	696,924516	768,875333	809,713548	784,822	734,809677
10:00	873,36793	888,283793	905,018387	904,177	855,89129	815,958333	817,931935	868,400968	938,068	968,795806	940,642	898,942581
11:00	981,71379	1008,56552	1013,67387	1004,617	947,258065	907,138333	913,085161	969,131935	1031,13867	1049,54226	1020,61567	991,612258
12:00	1013,66276	1026,69862	1041,15613	1026,17933	960,360645	937,602	953,915161	1016,53452	1053,18233	1051,04419	1015,68067	1009,26839
13:00	956,43966	975,387586	989,096774	962,941333	894,966839	880,428	901,568065	958,816129	983,033	968,470323	933,562	943,564839
14:00	836,98828	850,698621	855,78	819,917667	756,406129	752,069667	776,457097	823,623548	833,534333	814,169032	776,327	801,66
15:00	645,55897	657,534483	657,507097	801,010333	558,180645	563,498	588,986452	624,615484	618,755	584,313548	557,640333	596,387742
16:00	405,81517	414,19931	414,804194	361,494	323,737419	336,207333	360,353548	380,835806	361,672	323,542581	304,973667	351,242581
17:00	156,65207	164,337931	156,068065	116,763333	99,0654839	114,357667	127,378065	134,312581	108,345	74,9254839	69,7176667	111,00129
18:00	7,62069	10,5144828	6,95322581	0,37066667	0	0	3,97451613	2,4416129	0	0	0	0,24193548
19:00	0	0	0	0	0	0	0	0	0	0	0	0
20:00	0	0	0	0	0	0	0	0	0	0	0	0
21:00	0	0	0	0	0	0	0	0	0	0	0	0
22:00	0	0	0	0	0	0	0	0	0	0	0	0
23:00	0	0	0	0	0	0	0	0	0	0	0	0

ANEXO 24: Curva de irradiación promedio de cada mes.

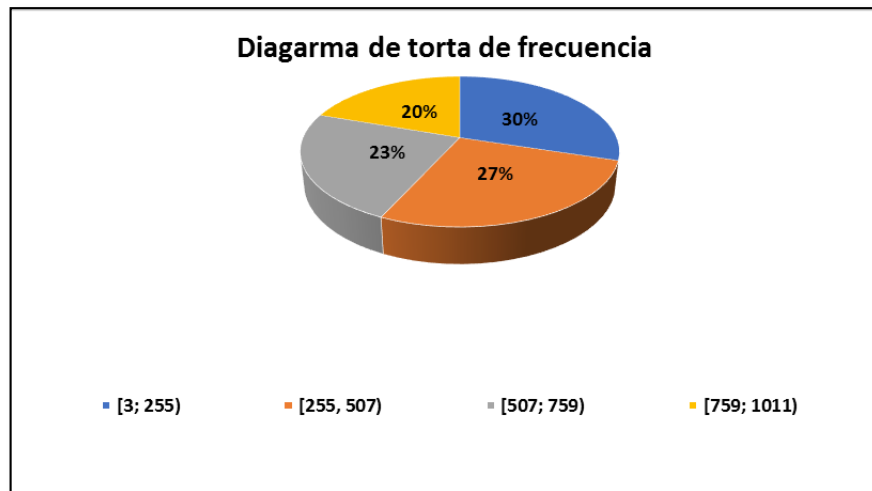




ANEXO 25: Tabla de frecuencias de irradiación solar.

	<i>Tiempos de Espera</i>	Frecuencia absoluta	Frecuencia relativa	Frecuencia relativa	Frecuencia porcentual	Frecuencia porcentual	Porción de la Torta
[3; 254]	<b>[3; 255)</b>	<b>3664</b>	0,29820135	0,30	29,8%	29,8%	107,35
[255; 506]	<b>[255, 507)</b>	<b>3322</b>	0,27036705	0,57	27,0%	56,9%	97,33
[507; 758]	<b>[507; 759)</b>	<b>2882</b>	0,23455685	0,80	23,5%	80,3%	84,44
[759; 1010]	<b>[759; 1011)</b>	<b>2419</b>	0,19687475	1,00	19,7%	100,0%	70,87
	<b>Total</b>	<b>12287</b>	<b>1</b>		<b>100%</b>		<b>360,00</b>

ANEXO 26: Torta de frecuencias de irradiación solar.

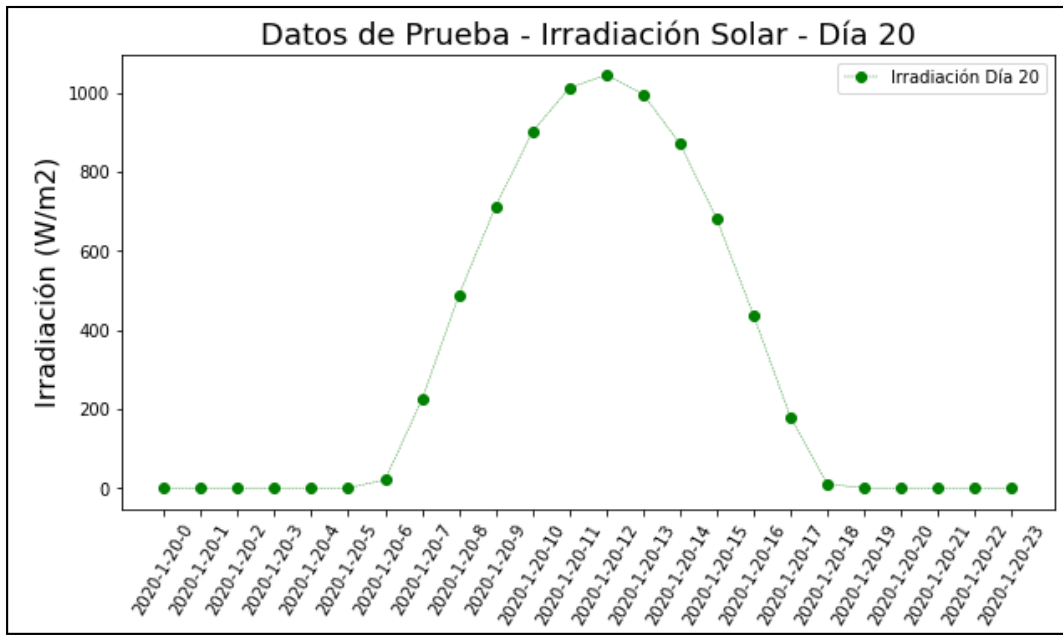


ANEXO 27: Tabla de prueba del modelo de predicción.

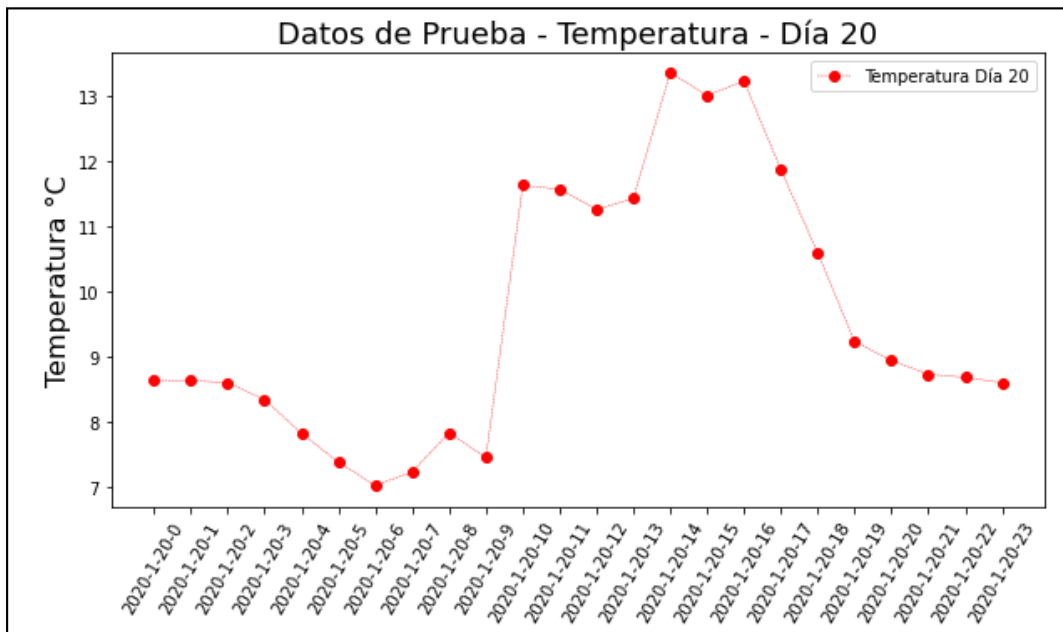
The shape of our features is: (24, 8)

	Year	Mes	Dia	Hora	Tiempo	Irradiacion	Temperatura	Potencia
0	2020	1	1	A	0	0.00	8.64	4.000000
1	2020	1	1	B	1	0.00	8.63	4.000000
2	2020	1	1	C	2	0.00	8.60	4.000000
3	2020	1	1	D	3	0.00	8.34	4.000000
4	2020	1	1	E	4	0.00	7.82	4.000000
5	2020	1	1	F	5	0.00	7.38	4.000000
6	2020	1	1	G	6	28.40	7.02	8.045584
7	2020	1	1	H	7	232.75	7.23	36.192254
8	2020	1	1	I	8	488.50	7.83	66.388250
9	2020	1	1	J	9	716.42	7.45	100.163758

ANEXO 28: Curva de Irradiación del día 20.



ANEXO 29: Curva de Temperatura del día 20.



### ANEXO 30: Código para graficar curvas.

```
# Gráfica Irradiación Solar MES 3
from datetime import datetime
import matplotlib.pyplot as plt
#plt.plot(fecha,features.iloc[:,5], 'g.-', alpha = 0.6)
plt.plot(fecha3,features3.iloc[:,5],color="green", lw=1, ls='--', label = 'Irradiación Mes Marzo')
plt.legend()
plt.gcf().set_size_inches(12, 6)
plt.xticks(rotation = '60');
plt.xticks(np.arange(0, 744, step=24))
plt.xlabel('Fecha (Año/Mes/Día/Hora)',fontsize=16, color="black")
plt.ylabel('Irradiación (W/m2)', fontsize=16, color="black")
plt.title(('Datos de Entrenamiento - Irradiación Solar - Mes Marzo'),fontsize=18, color="BLACK", alpha = 1)
plt.show()
print('#####')
# Gráfica Temperatura MES 3
#plt.plot(fecha,features.iloc[:,6], 'r.-', alpha = 1)
plt.plot(fecha3,features3.iloc[:,6],color="red", lw=0.8, ls='--', label = 'Temperatura Mes Marzo')
plt.legend()
plt.gcf().set_size_inches(12, 6)
plt.xticks(rotation = '60');
plt.xticks(np.arange(0, 744, step=24))
plt.xlabel('Fecha (Año/Mes/Día/Hora)',fontsize=16, color="black")
plt.ylabel('Temperatura °C', fontsize=16, color="black")
plt.title(('Datos de Entrenamiento - Temperatura - Mes Marzo'),fontsize=18, color="BLACK", alpha = 1)
plt.show()
print('#####')
```

### ANEXO 31: Pseudocódigo de importación de datos fotovoltaicos.

```
##### Proceso de importación de datos fotovoltaicos para PRUEBA
import pandas as pd # Librería PANDAS para importar datos
features5 = pd.read_csv('Datos_Solar_2020_mes5.csv') # Importación de datos fotovoltaicos
fecha5 = features5.iloc[:,0] # Fecha inicial para graficar
potencia_inicial5 = features5.iloc[:,8] # Potencia original base de datos
features5 = features5.drop(['Fecha'], axis=1) # Eliminación columna fecha de la matriz de datos
print('The shape of our features is:', features5.shape) # Verificación del tamaño de la matriz de datos
feature_list5 = list(features5.columns) # Guardo los tags de las columnas de la base de datos fotovoltaicos
features5.head(10) # Verificación de datos importados
```

ANEXO 32: Tabla de importación de datos fotovoltaicos mensuales.

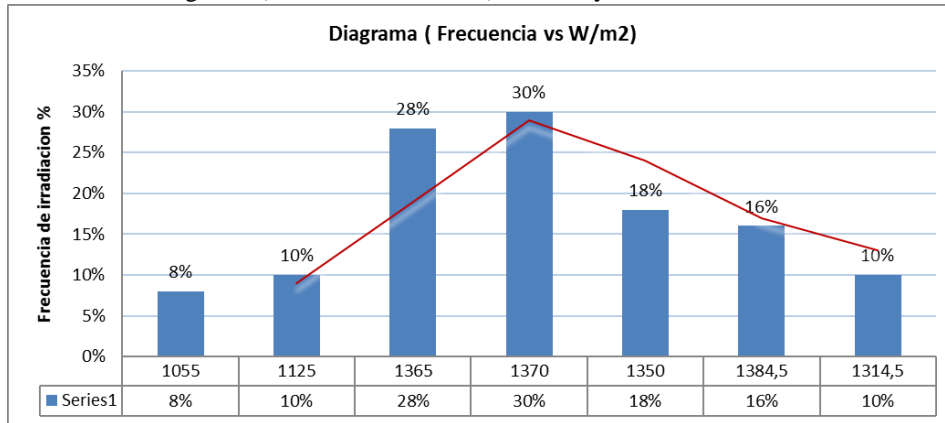
The shape of our features is: (744, 8)

	Year	Mes	Dia	Hora	Tiempo	Irradiacion	Temperatura	Potencia
0	2020	5	1	A	0	0.00	8.64	3.770000
1	2020	5	1	B	1	0.00	8.63	3.770000
2	2020	5	1	C	2	0.00	8.60	3.770000
3	2020	5	1	D	3	0.00	8.34	3.770000
4	2020	5	1	E	4	0.00	7.82	3.770000
5	2020	5	1	F	5	0.00	7.38	3.770000
6	2020	5	1	G	6	26.95	7.02	7.609031
7	2020	5	1	H	7	185.92	7.23	29.485076
8	2020	5	1	I	8	417.35	7.83	57.071405
9	2020	5	1	J	9	641.30	7.45	89.850537

ANEXO 33: Tabla diagrama de frecuencias.

INCIDENCIA DE IRRADIACIÓN 11:00 A 13:00 am					
Item	Medida Minima (W/m2)	Medida Maxima (W/m2)	Medida Promedio (W/m2)	Frecuencia de Incidencia	Procentaje de Incidencia
1	87	168	127,5	16	32,0%
2	165	235	200	8	16,0%
3	260	320	290	6	12,0%
4	280	297	288,5	9	18,0%
5	390	450	420	11	22,0%
6	380	420	400	7	14,0%
7	290	360	325	5	10,0%
N de Muestras		62			
Medida Minima (W/m2)		264,6			
Medida Míaxima (W/m2)		321,4			
Intervalo de medida (s)		80			

ANEXO 34: Diagrama (Frecuencia vs W/m2) de la mayor incidencia de irradiación solar.



ANEXO 35: Tabla diagrama de frecuencias de menor incidencia de irradiación solar.

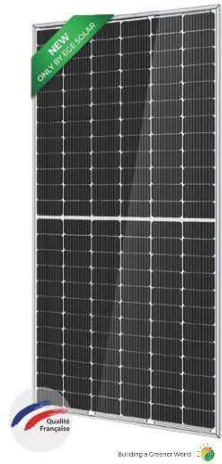
INCIDENCIA DE IRRADIACIÓN 15:00 A 17:00					
Item	Medida Mínima (W/m2)	Medida Máxima (W/m2)	Medida Promedio (W/m2)	Frecuencia de Incidencia	Porcentaje de Incidencia
1	830	1280	1055	4	8,0%
2	980	1270	1125	5	10,0%
3	1250	1480	1365	14	28,0%
4	1360	1380	1370	15	30,0%
5	1280	1420	1350	9	18,0%
6	1380	1389	1384,5	8	16,0%
7	1200	1429	1314,5	5	10,0%
N de Muestras		60			
Medida Mínima (W/m2)		1182,9			
Medida Máxima (W/m2)		1378,3			
Intervalo de medida (s)		80			

ANEXO 36: Panel fotovoltaico ECOGREE ENERGY.



# ATLAS

530-550 W / 144 Cells / 10 Bus Bars  
Monocrystalline Module  
EGE-530-550W-144M (M10)



Eco Green Energy's modules are only made of grade A solar cells with a very high efficiency and ensured more than 25 years lifespan.

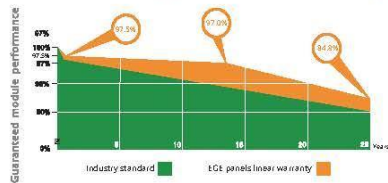
### KEY FEATURES

-  Lower LCOE and BOS
-  Anti PID/ Low LID protection
-  Less Hot Spot Shading effects
-  Lower temperature coefficient
-  More power output in weak light condition, such as haze, cloudy, and morning.

### CERTIFICATES

ISO 9001:2015 / PV CYCLE / CE

### LINEAR PERFORMANCE WARRANTY 12-Year Product Warranty - 25-Year Linear Power Warranty



**25 YEARS** OUTPUT POWER WARRANTY

**12 YEARS** PRODUCT WARRANTY



Eco Green Energy Group Ltd. 2018. All rights reserved.

## ANEXO 37: DataSheet del panel fotovoltaico ECOGREEN ENERGY.

### ELECTRICAL DATA AT STC\*

Power output (Pmax)	530 W	535 W	540 W	545 W	550 W
Power tolerance	0~+5 W	0~+5 W	0~+5 W	0~+5 W	0~+5 W
Module efficiency	20.7 %	20.9 %	21.1 %	21.3 %	21.5 %
Maximum power voltage (Vmp)	40.80 V	41.00 V	41.20 V	41.40 V	41.60 V
Maximum power current (Imp)	13.00 A	13.05 A	13.11 A	13.17 A	13.23 A
Open circuit voltage (Voc)	49.00 V	49.20 V	49.40 V	49.60 V	49.80 V
Short circuit current (Isc)	13.76 A	13.81 A	13.87 A	13.93 A	13.99 A

\*Standard Test Conditions: Irradiance: 1 000 W / m<sup>2</sup> • Cell temperature: 25°C • AM: 1.5

### ELECTRICAL DATA AT NMOT\*

Power output (Pmax)	395 W	399 W	402 W	406 W	410 W
Maximum power voltage (Vmp)	38.00 V	38.20 V	38.40 V	38.60 V	38.80 V
Maximum power current (Imp)	10.40 A	10.44 A	10.49 A	10.54 A	10.58 A
Open circuit voltage (Voc)	45.90 V	46.10 V	46.30 V	46.40 V	46.60 V
Short circuit current (Isc)	11.09 A	11.13 A	11.18 A	11.23 A	11.28 A

\*Nominal Operating Cell Temperature: Irradiance: 800 W / m<sup>2</sup> • Ambient temperature: 20°C • AM: 1.5 • Wind speed: 1 m/s

### TEMPERATURE CHARACTERISTICS

NMOT	41 °C ±3 °C
Temperature coefficient of Pmax	-0.36%/°C
Temperature coefficient of Voc	-0.28%/°C
Temperature coefficient of Isc	+0.05%/°C

### MAXIMUM RATINGS

Operating temperature range	-40 °C ~+85 °C
Maximum system voltage	1500V
Max series fuse rating	25 A
Max front load (e.g.: snow)	5400 Pa
Max back load (e.g.: wind)	2400 Pa

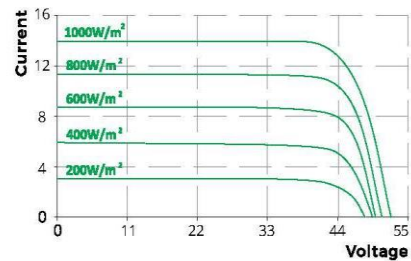
### MECHANICAL CHARACTERISTICS

Cell type	Monocrystalline (182x182 mm)
Number of cells	144
Dimensions	2279x1134x35mm (1.5mm Cell Gap)
Weight	29 kg
Glass	3.2 mm tempered glass
Frame	Anodized aluminium alloy
Junction box	IP68
Cable	4.0 mm <sup>2</sup> Landscape: N 1400mm/P 1400mm Portrait: N 200mm/P 300mm
Connector	MC4 compatible

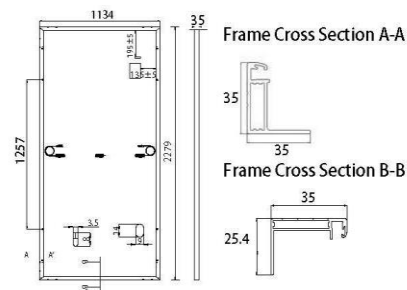
### PACKAGING (2279x1134x35mm)

Type	Pcs/ Type	Weight
Per Pallet	30 pcs	910 kg
20ft GP Container	240pcs (5pallets+45cartons)	7.28 t
40ft GP Container	510PCS (10pallets+105cartons)	15.47 t
40ft HQ Container	600PCS (20pallets)	18.20 t

### I-V CURVES



### ENGINEERING DRAWINGS (mm)



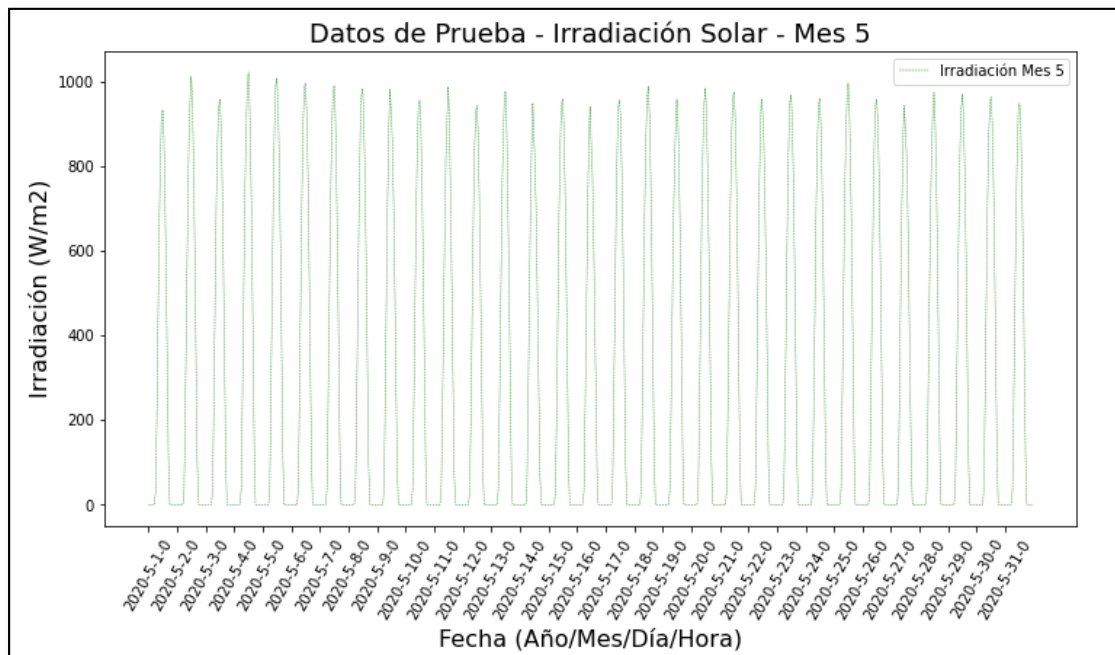
Eco Green Energy Group Ltd, 2018. All rights reserved.  
Add: 299 Xing Cheng Road, Chong Chuan District, Nantong,  
Jiangsu, China  
Tel: +86 513 66690088 / E-mail: info@eco-greenenergy.com

Specifications included in this datasheet are subject to change  
without prior notice. Refer to our website for further information  
or contact one of our sales staff.  
www.eco-greenenergy.com

ANEXO 38: Tabla de variables utilizadas.

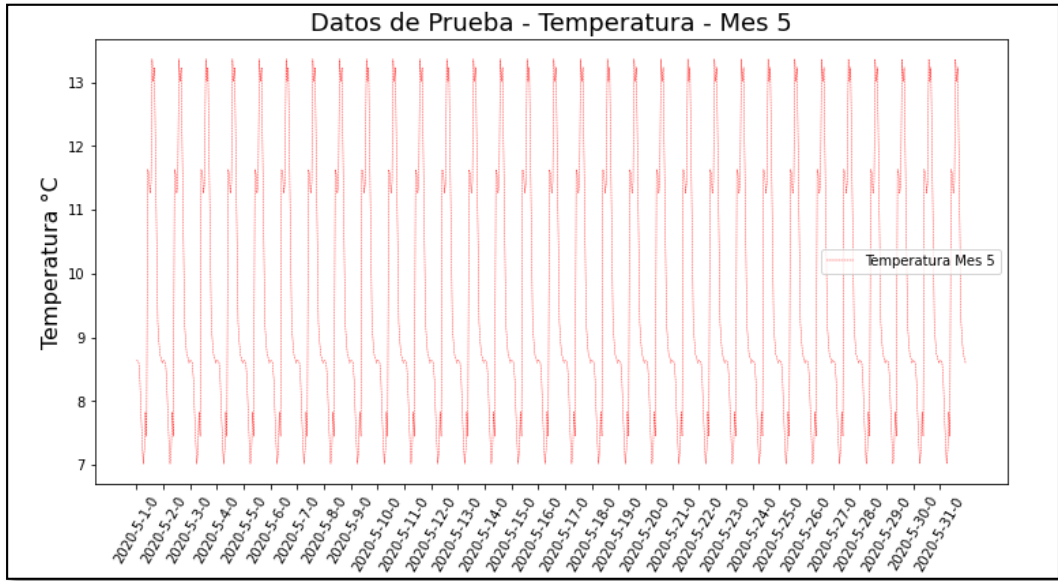
Variable: Irradiacion	Importance: 0.95
Variable: Temperatura	Importance: 0.02
Variable: Tiempo	Importance: 0.01
Variable: Year	Importance: 0.0
Variable: Mes	Importance: 0.0
Variable: Dia	Importance: 0.0
Variable: Hora_A	Importance: 0.0
Variable: Hora_B	Importance: 0.0
Variable: Hora_C	Importance: 0.0
Variable: Hora_D	Importance: 0.0
Variable: Hora_E	Importance: 0.0
Variable: Hora_F	Importance: 0.0
Variable: Hora_G	Importance: 0.0
Variable: Hora_H	Importance: 0.0
Variable: Hora_I	Importance: 0.0
Variable: Hora_J	Importance: 0.0
Variable: Hora_K	Importance: 0.0
Variable: Hora_L	Importance: 0.0
Variable: Hora_M	Importance: 0.0
Variable: Hora_N	Importance: 0.0
Variable: Hora_O	Importance: 0.0
Variable: Hora_P	Importance: 0.0
Variable: Hora_Q	Importance: 0.0
Variable: Hora_R	Importance: 0.0
Variable: Hora_S	Importance: 0.0
Variable: Hora_T	Importance: 0.0
Variable: Hora_U	Importance: 0.0
Variable: Hora_V	Importance: 0.0
Variable: Hora_W	Importance: 0.0
Variable: Hora_X	Importance: 0.0

ANEXO 39: Curva de Irradiación solar mes de mayo.

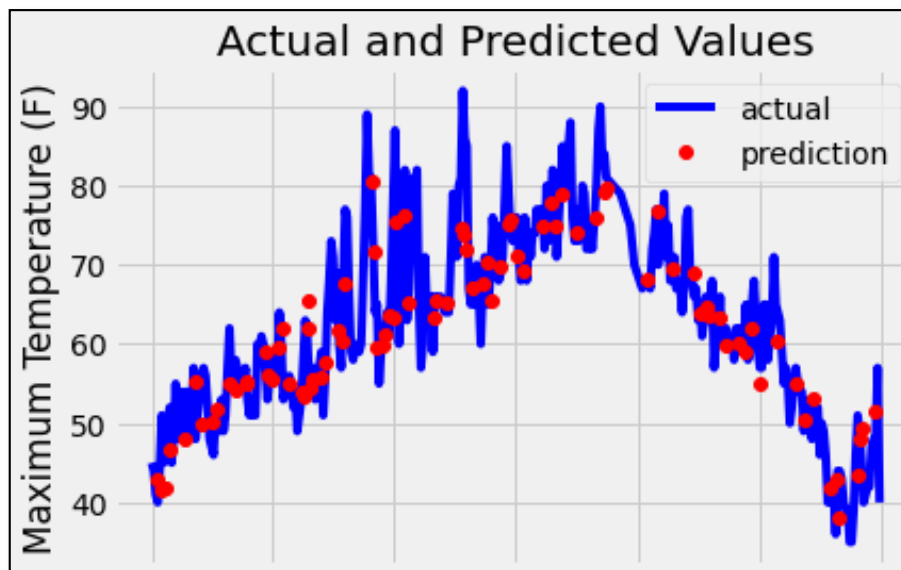




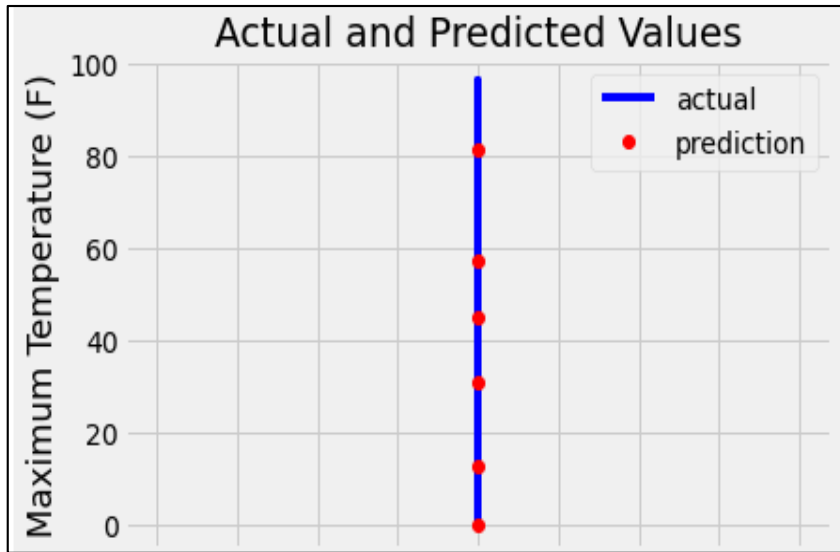
ANEXO 40: Curva de Temperatura mes de Mayo



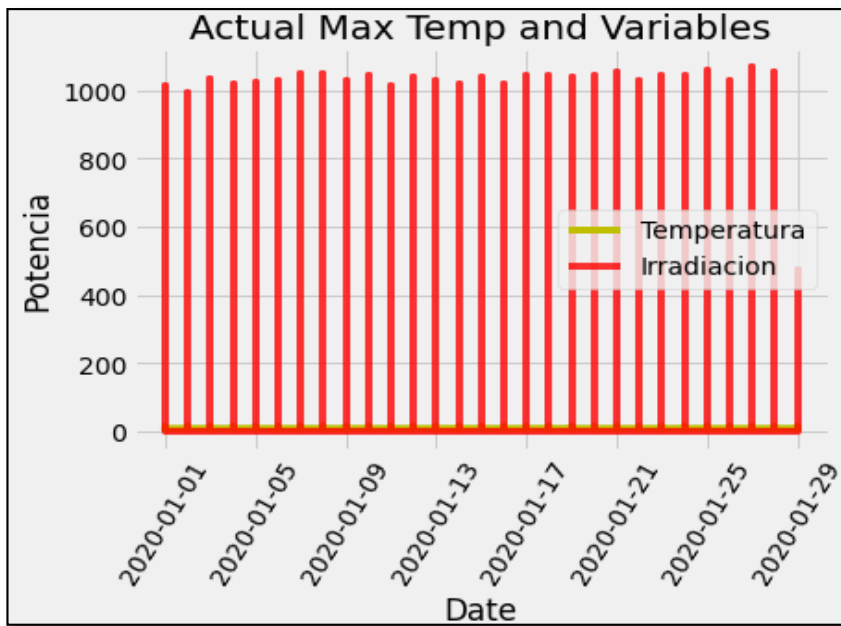
ANEXO 41: Pruebas de predicción.



ANEXO 42: Curva de Temperatura mes de mayo.



ANEXO 43: Curva de Temperatura mes de Mayo.



#### ANEXO 44: Cálculo de la potencia.

Mean Absolute Error: 2.37 Watios.

[7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 3.36929050e+00 2.89689223e-01  
1.13762254e+01 1.77422802e+01 1.08655675e+01 6.79050954e+00  
1.05504929e+01 1.11789275e+01 1.29671828e+01 1.62204762e+01  
3.84935834e+01 1.92521046e+01 1.04505276e+01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 9.90056997e+00 1.92582150e-01  
2.05142043e+00 2.18761385e+01 5.10743697e+00 3.04551393e+00  
2.31201864e+00 2.54023426e+00 3.49935268e+00 6.33395262e+00  
1.56187521e+01 1.70098186e+01 1.04480229e+01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 1.52087284e+01 6.09539229e-01  
4.66336488e-01 2.35215369e+01 5.60515664e+00 4.60234245e+00  
7.71087461e+00 7.33381106e+00 7.42173254e+00 1.20801211e+01  
2.28211330e+01 1.99280244e+01 1.04474969e+01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 1.55072221e+01 7.21135744e-01  
7.86015095e-02 5.68108514e-01 2.69092970e+00 4.16208872e+00  
1.42183003e+00 1.41691533e+00 1.91209191e+00 3.78597968e+00  
1.11219947e+01 7.62914901e+00 1.04426377e+01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 2.16786311e+01 4.09846342e+00  
1.81404081e-01 4.67212477e-02 5.55610657e-01 2.51862338e+00  
2.65558410e+00 2.57454287e+00 3.07954189e+00 5.51704986e+00  
1.39789526e+01 1.36908440e+01 1.04377034e+01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 2.08067820e+01 3.00222999e+00  
1.89928022e-01 1.57803713e-02 1.89593587e-01 2.66758264e+00  
3.87313087e+00 4.09684908e+00 4.71486979e+00 7.20904151e+00  
1.60303675e+01 1.69790609e+01 1.04015349e+01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 1.46164397e+01 4.65011245e-01  
2.69692728e-01 1.30695316e+01 3.20188365e+00 1.70655723e+00  
4.41507569e+00 4.31356537e+00 5.47531284e+00 9.21133354e+00  
2.02853181e+01 2.11202014e+01 1.02862668e+01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 1.82566540e+01 2.31664366e+00  
7.17279664e-02 6.55520416e-01 1.30686997e+00 3.40749999e+00  
1.13003355e+01 1.18916306e+01 1.36803613e+01 1.52548705e+01  
3.46638270e+01 2.82577761e+01 1.01222063e+01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 1.21280236e+01 7.61058062e-02  
9.38290062e-02 2.29800608e+01 5.23251823e+00 3.09214738e+00  
8.03646469e+00 9.31170775e+00 1.07479458e+01 1.48372029e+01  
3.10323014e+01 2.40472591e+01 1.00779225e+01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 5.04559324e+00 1.52510578e-01  
1.44392020e+01 1.86866855e+01 9.14773898e+00 6.92711382e+00

4.68409648e+00 8.31481343e+00 1.39014735e+01 1.37102953e+01  
4.93555422e+01 1.95602834e+00 1.00592371e+01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 5.48410088e+00 2.07706008e-02  
1.60373679e+01 2.06044797e+01 7.85109968e+00 5.51499758e+00  
9.52084851e+00 1.01345628e+01 1.15656481e+01 1.51292728e+01  
3.39668606e+01 2.68596401e+01 1.00332880e+01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 1.29561789e+01 3.89121437e-01  
9.20647738e-02 2.33123816e+01 4.61982703e+00 2.87833237e+00  
5.75646312e+00 7.88467879e+00 1.16513184e+01 1.68406283e+01  
3.89444819e+01 2.18571636e+01 9.95854655e+00 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 1.02773386e+01 3.67501055e-01  
8.75550102e-01 2.23982456e+01 5.51220383e+00 3.16829963e+00  
2.17871646e+01 2.14007162e+01 2.42194808e+00 1.55593129e+01  
5.46063550e+01 3.03398663e+00 9.46616472e+00 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 6.41535075e+00 4.42880457e-01  
1.43372454e+00 2.20697287e+01 6.05073364e+00 3.74505423e+00  
7.66950040e+00 7.79428038e+00 8.51372267e+00 1.25980015e+01  
2.48162325e+01 2.18763916e+01 9.33281235e+00 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 1.09234898e+01 6.71686737e-03  
2.33103076e-01 2.24028240e+01 5.73878312e+00 3.99786650e+00  
2.27624987e+01 2.15292147e+01 3.98987708e+00 7.40741397e+00  
5.92627171e+01 4.32192263e+00 9.30931788e+00 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 1.33917985e+01 1.17836273e-01  
5.41327943e-02 2.33810753e+01 5.26751913e+00 3.59564938e+00  
7.98000054e+00 8.19482227e+00 9.15537350e+00 1.29452848e+01  
2.75593129e+01 1.95460440e+01 8.98097149e+00 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 1.47652767e+01 2.36062823e-01  
4.54043399e-01 1.27542709e+01 3.95110374e+00 2.48878916e+00  
4.54877104e+00 4.82067472e+00 5.51766110e+00 9.26518420e+00  
2.16455959e+01 2.12338517e+01 8.93112716e+00 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 1.56891723e+00 1.01234011e-01  
1.36221867e+01 1.74276797e+01 1.24549600e+01 9.84100640e+00  
7.71223755e+00 6.58419899e+00 5.62931545e+00 8.46018290e+00  
1.67120377e+01 1.53772774e+01 8.87857771e+00 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 9.24511577e+00 9.48510841e-01  
7.60994009e-02 2.34051003e+01 4.78628509e+00 3.23947985e+00  
5.01314586e+00 5.36915316e+00 5.83905551e+00 8.83052822e+00  
1.83926049e+01 1.40088012e+01 8.86109462e+00 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 1.45276427e+01 2.44855775e+00  
3.02287026e-01 4.73095042e-01 3.43527821e+00 1.99146626e+00  
5.95976286e+00 5.83752463e+00 6.26379800e+00 9.46973981e+00  
1.88458849e+01 1.75912278e+01 8.85257850e+00 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
7.95755968e-01 7.95755968e-01 1.25169395e+01 4.47950594e-01

2.05688682e-01 1.30297399e+01 3.51032419e+00 2.11666050e+00  
 1.08290595e+01 1.13333507e+01 1.19784696e+01 1.50820770e+01  
 3.16468080e+01 2.12881883e+01 8.83532084e+00 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 1.04866155e+01 8.57482526e-01  
 6.47805867e-02 2.31950581e+01 4.72291536e+00 2.90144836e+00  
 6.68812185e+00 7.03187454e+00 8.37102470e+00 1.26128837e+01  
 2.74735477e+01 2.08868942e+01 8.69853698e+00 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 9.47195238e+00 9.15847558e-01  
 7.50393172e-02 2.26185509e+01 5.36704445e+00 3.55910925e+00  
 7.49351951e+00 7.24145130e+00 7.93076346e+00 1.14027203e+01  
 2.23747591e+01 1.96260462e+01 8.60067678e+00 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 1.47762443e+01 2.85641352e+00  
 4.12065959e-01 4.64916471e-01 3.51851458e+00 1.94685535e+00  
 3.80049361e+00 3.72164162e+00 4.30420910e+00 5.90173940e+00  
 1.24786228e+01 9.94091159e+00 8.51541542e+00 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 1.42051566e+01 1.48640771e+00  
 1.50720145e-01 1.25789913e+01 4.94455317e+00 3.39987050e+00  
 7.73715513e+00 7.67647502e+00 7.89993732e+00 1.08383433e+01  
 1.78370636e+01 1.56041800e+01 8.23355700e+00 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 1.13464879e+01 9.21843714e-01  
 3.50397335e-01 2.32101514e+01 5.24172661e+00 3.62253861e+00  
 1.88432751e+01 1.63800416e+01 1.44755162e+00 1.62226328e+01  
 3.43098798e+01 1.67045016e+01 7.92807888e+00 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.58104001e+00 5.02835138e-01  
 6.14702288e+00 2.07989850e+01 8.17020752e+00 6.10805770e+00  
 5.97767926e+00 5.73067686e+00 5.37625428e+00 6.82756325e+00  
 9.94564372e+00 7.69894108e+00 7.86886683e+00 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 1.24613921e+01 7.95945136e+00  
 1.16146916e-01 2.31812650e+01 5.91426955e+00 3.21859086e+00  
 5.34961785e+00 4.92457803e+00 3.78478588e+00 3.52984561e+00  
 2.93839664e+00 3.87896548e+00 7.67484965e+00 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 1.64987850e+01 2.45983757e+01  
 3.53695498e-01 1.29118543e+01 4.30392175e+00 1.99195543e+00  
 6.13356747e+00 5.76544686e+00 4.84241250e+00 6.71455481e+00  
 2.00619584e+00 3.52356837e+00 7.67484965e+00 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 1.03819047e+01 2.00081106e+01  
 1.96155020e-02 2.25657850e+01 6.62506741e+00 3.69008260e+00  
 7.68117866e+00 7.51484678e+00 6.70275048e+00 8.03742989e+00  
 2.01192412e+00 1.13532459e+01 7.67484965e+00 7.95755968e-01  
 7.95755968e-01 7.95755968e-01 7.95755968e-01 7.95755968e-01]

Accuracy: 94.47 %.

ANEXO 45: Resultados arrojados por el programa.

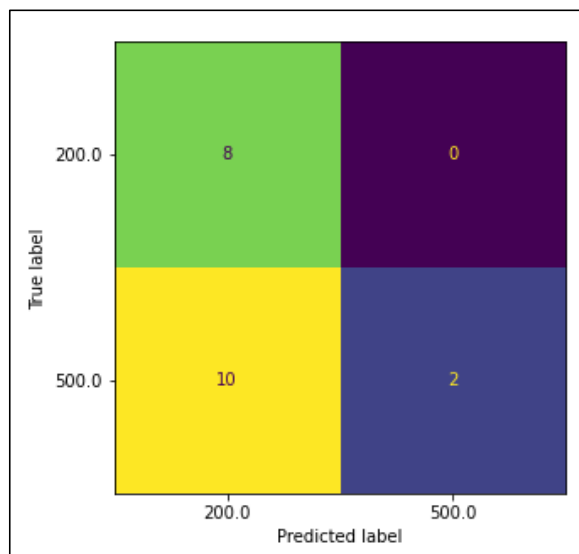
```
ALGORITMO K-MEANS
from sklearn.cluster import KMeans
X = datos[:,0:-1]
data

Precisión promedio: 0.7222222222222222
Exactitud promedio: 0.5833333333333334
Medida-F1 promedio: 0.4505494505494505

ALGORITMO REDES NEURONALES
Precisión promedio: 0.478021978021978
Exactitud promedio: 0.47916666666666663
Medida-F1 promedio: 0.4486215538847118

ALGORITMO REDES NEURONALES
Precisión promedio: 0.478021978021978
Exactitud promedio: 0.47916666666666663
Medida-F1 promedio: 0.4486215538847118
```

ANEXO 46: Matriz de confusión con k-means.



ANEXO 47: Matriz de confusión con Redes neuronales.

