

	GESTIÓN DE SERVICIOS ACADÉMICOS Y BIBLIOTECARIOS	CÓDIGO	FO-GS-15	
		VERSIÓN	02	
	ESQUEMA HOJA DE RESUMEN		FECHA	03/04/2017
			PÁGINA	1 de 1
ELABORÓ	REVISÓ	APROBÓ		
Jefe División de Biblioteca	Equipo Operativo de Calidad	Líder de Calidad		

RESUMEN TRABAJO DE GRADO

AUTORES:

NOMBRES: GIAN FRANCO

APELLIDOS: DI CANDIA JAIMES

NOMBRES: BRAYAN ORLANDO

APELLIDOS: CALDERÓN ANGARITA

FACULTAD: INGENIERÍA

PLAN DE ESTUDIOS: INGENIERÍA ELECTRÓNICA

DIRECTORES:

NOMBRE(S): SERGIO

APELLIDOS: CASTRO CASADIEGO

NOMBRES(S): BYRON

APELLIDOS: MEDINA DELGADO

TÍTULO DEL TRABAJO (TESIS): ANÁLISIS DEL RECURSO EÓLICO EN LA CIUDAD DE CÚCUTA MEDIANTE CIENCIA DE LOS DATOS E INTELIGENCIA ARTIFICIAL.

RESUMEN:

En la actualidad el uso de energías renovables es cada vez más utilizadas en el mundo, debido a que no afectan negativamente al medio ambiente, la energía del viento es de las más utilizadas también es conocida como Recurso Eólico. El presente proyecto tuvo como objetivo la predicción del Recurso Eólico disponible promedio y total en la ciudad de Cúcuta, con la utilización de las dos variables más influyentes velocidad del viento y temperatura. Para la construcción del modelo se utilizó el enfoque de Aprendizaje Automático en conjunto con la técnica de Redes Neuronales Artificiales. La métrica de éxito utilizada para evaluar al modelo fue el Error Absoluto Medio (MAE). El cual para la etapa de entrenamiento con un rango de 1 año tuvo un valor de 1.6 el cual representa el 1.7% del valor promedio de las predicciones. Para la etapa de evaluación del modelo se alimentó el algoritmo con datos tomados experimentalmente en las zonas de Ceiba, Torcoroma y Belén. Para evaluar predicciones en cada una de estas zonas con un rango de 10 minutos se obtuvo un MAE de 0.02, 0.03 y 0.06 respectivamente.

PALABRAS CLAVES: RECURSO EÓLICO, APRENDIZAJE AUTOMÁTICO, REDES NEURONALES ARTIFICIALES, MÉTRICA DE ÉXITO, ERROR ABSOLUTO MEDIO

CARACTERÍSTICAS:

PÁGINAS: 162

PLANOS: 0

ILUSTRACIONES: 48

CD ROOM: 0

ANÁLISIS DEL RECURSO EÓLICO EN LA CIUDAD DE CÚCUTA MEDIANTE
CIENCIA DE LOS DATOS E INTELIGENCIA ARTIFICIAL

GIAN FRANCO DI CANDIA JAIMES
BRAYAN ORLANDO CALDERÓN ANGARITA

UNIVERSIDAD FRANCISCO DE PAULA SANTANDER
FACULTAD DE INGENIERÍA
PLAN DE ESTUDIOS DE INGENIERÍA ELECTRÓNICA
SAN JOSÉ DE CÚCUTA

2022

ANÁLISIS DEL RECURSO EÓLICO EN LA CIUDAD DE CÚCUTA MEDIANTE
CIENCIA DE LOS DATOS E INTELIGENCIA ARTIFICIAL

GIAN FRANCO DI CANDIA JAIMES
BRAYAN ORLANDO CALDERÓN ANGARITA

Director:

I.E. MSc. Sergio Castro Casadiego

Codirector:

Ph.D. Byron Medina Delgado

UNIVERSIDAD FRANCISCO DE PAULA SANTANDER

FACULTAD DE INGENIERÍA

PLAN DE ESTUDIOS DE INGENIERÍA ELECTRÓNICA

SAN JOSÉ DE CÚCUTA

2022

ACTA DE SUSTENTACIÓN DE UN TRABAJO DE GRADO

Ciudad y Fecha: CÚCUTA, 15 DE MARZO DE 2022

Hora: 15:00

Lugar: SALON SC 301

Plan de Estudios: INGENIERÍA ELECTRÓNICA

Título de la Tesis: “ANÁLISIS DEL RECURSO EÓLICO EN LA CIUDAD DE CÚCUTA MEDIANTE CIENCIA DE LOS DATOS E INTELIGENCIA ARTIFICIAL”

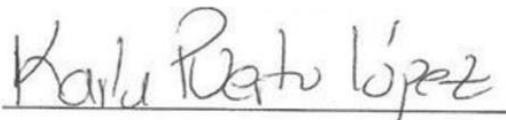
Jurados: KARLA CECILIA PUERTO LÓPEZ, IE MSc
ORIANA ALEXANDRA LÓPEZ BUSTAMANTE, IE MSc

Director: SERGIO ALEXANDER CASTRO CASADIEGO, IE MSc

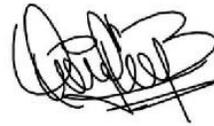
Codirector: BYRON MEDINA DELGADO, IE PhD

Nombre del Estudiante:	Código:	Calificación:	
		Número	Letra
GIAN FRANCO DI CANDIA JAIMES	1161520	5,0	Cinco, cero

LAUREADO



KARLA CECILIA PUERTO LÓPEZ



ORIANA ALEXANDRA LÓPEZ BUSTAMANTE



ANGELO JOSEPH SOTO VERGEL
Coordinador Comité Curricular
Ingeniería Electrónica

ACTA DE SUSTENTACIÓN DE UN TRABAJO DE GRADO

Ciudad y Fecha: CÚCUTA, 15 DE MARZO DE 2022

Hora: 15:00

Lugar: SALON SC 301

Plan de Estudios: INGENIERÍA ELECTRÓNICA

Título de la Tesis: “ANÁLISIS DEL RECURSO EÓLICO EN LA CIUDAD DE CÚCUTA MEDIANTE CIENCIA DE LOS DATOS E INTELIGENCIA ARTIFICIAL”

Jurados: KARLA CECILIA PUERTO LÓPEZ, IE MSc
ORIANA ALEXANDRA LÓPEZ BUSTAMANTE, IE MSc

Director: SERGIO ALEXANDER CASTRO CASADIEGO, IE MSc

Codirector: BYRON MEDINA DELGADO, IE PhD

Nombre del Estudiante:	Código:	Calificación:	
		Número	Letra
BRAYAN ORLANDO CALDERÓN ANGARITA	1161162	5,0	Cinco, cero

LAUREADO


KARLA CECILIA PUERTO LÓPEZ


ORIANA ALEXANDRA LÓPEZ BUSTAMANTE


ANGELO JOSEPH SOTO VERGEL
Coordinador Comité Curricular
Ingeniería Electrónica

Tabla de contenido

1. Introducción	16
2. Descripción del problema	17
2.1. Planteamiento del problema	17
3. Justificación	19
3.1. Beneficios	19
3.1.1. Beneficios Ambientales	20
3.1.2. Beneficios Sociales	20
3.1.3. Beneficios Institucionales	20
3.1.4. Beneficios Económicos	21
4. Alcance	22
5. Limitaciones y delimitaciones	23
5.1. Limitaciones	23
5.2. Delimitaciones	23
5.2.1. Espacio geográfico y tiempo	23
5.2.2. Variables	24
5.2.3. Profundidad del tema	24
6. Objetivos	25
6.1. General	25
6.2. Específicos	25

7. Marco referencial	26
7.1. Antecedentes	26
7.2. Marco Teórico	29
7.2.1. Recurso Eólico	29
7.2.2. Variación de la velocidad media a largo plazo con la altura	31
7.2.3. Variación de la velocidad del viento estacionario con la altura	31
7.2.4. Inteligencia Artificial y Aprendizaje Automático	31
7.2.5. Aprendizaje Profundo	33
7.2.6. Ciencia de los Datos	33
7.2.7. Redes Neuronales Artificiales	34
7.2.8. Perceptrón simple	35
7.2.9. Perceptrón multicapa	36
7.2.10. Retropropagación	37
7.2.11. Validación K-Fold	38
7.2.12. Python	38
7.2.13. Anemómetro de álabe	38
7.3. Marco Legal	39
8. Diseño metodológico	42
8.1. Objetivo 1. Recopilar información sobre Recurso Eólico e Inteligencia Artificial	42
8.2. Objetivo 2. Seleccionar la tecnología y herramienta de software para la Inteligencia	

Artificial	42
8.3. Objetivo 3. Creación del modelo de Ciencia de los Datos e Inteligencia Artificial	43
8.4. Objetivo 4. Evaluar los resultados para determinar las zonas estratégicas donde se pueda aprovechar el recurso eólico en la ciudad de Cúcuta	44
9. Resultados	45
9.1. Estado del arte sobre Recurso Eólico e Inteligencia Artificial	45
9.1.1. Bases de datos para el Recurso Eólico	46
9.1.2. Selección del gestor de referencia	46
9.2. Selección de técnicas y herramientas de software	50
9.2.1. Técnicas de Inteligencia Artificial	51
9.2.2. Selección del lenguaje de programación	56
9.2.3. Software y entornos virtuales	59
9.3. Creación del modelo de Ciencia de los Datos	60
9.3.1. Colección de datos	61
9.3.2. Preprocesamiento de los datos	63
9.3.3. Exploración de los datos	65
9.3.4. Entrenamiento del modelo	68
9.3.5. Parámetros	68
9.3.6. Evaluación del modelo	70
9.3.7. Uso del modelo	71

9.3.8. Pruebas de funcionamiento sobre datos experimentales	71
9.3.9. Presentación de resultados mediante interfaz gráfica	85
9.4. Selección de zona para mejor aprovechamiento del Recurso Eólico	86
9.4.1. Análisis del perfil vertical del viento	86
9.4.2. Análisis del Recurso Eólico disponible promedio y total	87
10. Conclusiones	92
11. Recomendaciones	94
Referencias	96
Anexos	101

Lista de figuras

Fig. 1. Tipos de Aprendizaje Automático.	33
Fig. 2. Estructura jerárquica de un sistema basado en RNA.	35
Fig. 3. Arquitectura (izquierda) y función de transferencia (derecha) de un perceptrón simple.	36
Fig. 4. Región de decisión correspondiente a un perceptrón simple.	36
Fig. 5 Regiones de decisión obtenidas para el perceptrón simple (arriba), el perceptrón multicapa con una capa oculta (en medio) y el perceptrón multicapa con dos capas ocultas (abajo).	37
Fig. 6. Arquitectura (izquierda) y función de activación (derecha) para el perceptrón multicapa.	37
Fig. 7. Anemómetro de álabes.	39
Fig. 8. Representación gráfica de los resultados de la selección del gestor de referencia.	50
Fig. 9. Representación gráfica de los resultados de la selección de la técnica de Inteligencia Artificial.	55
Fig. 10. Representación gráfica de los resultados de la selección del lenguaje de programación.	59
Fig. 11. Proceso de creación de un modelo de Aprendizaje Automático.	61
Fig. 12. Bases de datos cargadas y definidas mediante un marco de datos en Python.	62
Fig. 13. Marco de datos correspondiente a los valores de velocidad de viento para un rango de 4 años y una frecuencia de 2 minutos.	62
Fig. 14. Marco de datos correspondiente a los valores de velocidad de viento para un rango de 4 años y una frecuencia de 2 minutos.	63
Fig. 15. Conversión del tipo de dato datetime a numérico para la variable Fecha en el marco de	

datos de velocidad.	64
Fig. 16. Conversión del tipo de dato objeto a numérico para velocidad de viento en el marco de datos de velocidad.	64
Fig. 17. Conversión del tipo de dato datetime a numérico para la variable Fecha en el marco de datos de temperatura.	64
Fig. 18. Marco de datos de tipo numérico con las entradas del modelo.	65
Fig. 19. Resumen estadístico del marco de datos.	66
Fig. 20. Acote del rango y resumen estadístico del nuevo marco de datos.	66
Fig. 21. Conteo de valores nulos por columnas.	66
Fig. 22. Perfil de viento para la zona del Aeropuerto.	67
Fig. 23. Marco de datos con las entradas y salida del modelo.	68
Fig. 24. Resumen estadístico del marco de datos.	68
Fig. 25. Relación entre el número de épocas y el MAE para la zona de Aeropuerto.	71
Fig. 26. Valores reales vs valores predichos mensuales para el año 2020 zona Aeropuerto.	71
Fig. 27. Anemómetro fijado en la zona de Torcoroma.	73
Fig. 28. Marco de datos en Python para la zona de Ceiba.	73
Fig. 29. Marco de datos en Python para la zona de Torcoroma.	74
Fig. 30. Marco de datos en Python para la zona de Belén.	75
Fig. 31. Marco de datos de tipo numérico con las entradas del modelo para Ceiba.	76
Fig. 32. Marco de datos de tipo numérico con las entradas del modelo para Torcoroma.	77

Fig. 33. Marco de datos de tipo numérico con las entradas del modelo para Belén.	77
Fig. 34. Perfil vertical del viento para la zona 2 Ceiba.	78
Fig. 35. Perfil vertical del viento para la zona 3 Torcoroma.	79
Fig. 36. Perfil vertical del viento para la zona 4 Belén.	79
Fig. 37. Resumen estadístico del marco de datos para Ceiba.	80
Fig. 38. Resumen estadístico del marco de datos para Torcoroma.	80
Fig. 39. Resumen estadístico del marco de datos para Belén.	81
Fig. 40. Relación de Épocas vs MAE para zona 2 Ceiba.	83
Fig. 41. Relación de Épocas vs MAE para zona 3 Torcoroma.	83
Fig. 42. Relación de épocas vs MAE para zona 4 Belén.	83
Fig. 43. Valores reales y predichos para una predicción de 10 minutos en zona 2 Ceiba.	84
Fig. 44. Valores reales y predichos para una predicción de 10 minutos en zona 3 Torcoroma.	84
Fig. 45. Valores reales y predichos para una predicción de 10 minutos en zona 4 Belén.	85
Fig. 46. Interfaz gráfica con los resultados del modelo.	85
Fig. 47. Resultados del modelo para la zona 1.	86
Fig. 48. Comparación del perfil vertical del viento de todas las zonas.	87

Lista de tablas

Tabla 1. Asignación de puntaje.	46
Tabla 2. Parámetros de comparación del gestor de referencia.	47
Tabla 3. Peso de cada parámetro del gestor de referencia.	47
Tabla 4. Comparación del gestor de referencia respecto al coste (A).	48
Tabla 5. Comparación del gestor de referencia respecto al sistema operativo (B).	48
Tabla 6. Comparación del gestor de referencia respecto al almacenamiento (C).	48
Tabla 7. Comparación del gestor de referencia respecto al editor de formatos (D).	49
Tabla 8. Comparación del gestor de referencia respecto al plugin para Word (E).	49
Tabla 9. Matriz de decisión de la selección del gestor de referencia.	49
Tabla 10. Parámetros de comparación de la técnica de Inteligencia Artificial.	52
Tabla 11. Peso de cada parámetro de la técnica de Inteligencia Artificial.	52
Tabla 12. Comparación de las técnicas de Inteligencia Artificial respecto a las aplicaciones de predicción de tiempo (A).	52
Tabla 13. Comparación de las técnicas de Inteligencia Artificial respecto a la habilidad con funciones no lineales (B).	53
Tabla 14. Comparación de las técnicas de Inteligencia Artificial respecto a cotas de error (C).	53
Tabla 15. Comparación de las técnicas de Inteligencia Artificial respecto a la capacidad de entrenamiento (D).	53
Tabla 16. Matriz de decisión de la selección de la técnica de Inteligencia Artificial.	54
Tabla 17. Parámetros de comparación del lenguaje de programación.	56

Tabla 18. Peso de cada parámetro del lenguaje de programación.	56
Tabla 19. Comparación del lenguaje de programación respecto al sistema operativo (A).	57
Tabla 20. Comparación del lenguaje de programación respecto al tipo de licencia (B).	57
Tabla 21. Comparación del lenguaje de programación respecto a la disponibilidad de librerías (C).	57
Tabla 22. Comparación del lenguaje de programación respecto a la especialización del software (D).	58
Tabla 23. Comparación del lenguaje de programación respecto a la interpretación en entornos de desarrollo (E).	58
Tabla 24. Matriz de decisión de la selección del lenguaje de programación.	58
Tabla 25. Resumen de los parámetros utilizados para el entrenamiento del modelo.	69
Tabla 26. Relación entre el número de épocas y el MAE.	70
Tabla 27. Valores MAE para distintas épocas en cada zona.	82
Tabla 28. Recurso Eólico disponible promedio y total para la zona del Aeropuerto.	87
Tabla 29. Recurso Eólico disponible promedio y total para la zona de Belén.	89
Tabla 30. Recurso Eólico disponible promedio y total para la zona de Ceiba.	89
Tabla 31. Recurso Eólico disponible promedio y total para la zona de Torcoroma.	90

Lista de anexos

Anexo 1. Ponencia: Análisis del Recurso Eólico disponible en la ciudad de Cúcuta usando aprendizaje automático	101
Anexo 2. Certificado de participación en VI Congreso de Investigación, Desarrollo e Innovación de la Universidad Internacional de Ciencia y Tecnología IDI-UNICyT 2021	122
Anexo 3. Exposición del proyecto en el VI Congreso de Investigación, Desarrollo e Innovación de la Universidad Internacional de Ciencia y Tecnología IDI-UNICyT 2021	124
Anexo 4. Artículo: Predicción del recurso eólico disponible en la ciudad de Cúcuta usando aprendizaje automático	126
Anexo 5. Divulgación en el semillero SINE mediante la plataforma Google Meet	151
Anexo 6. Manual de usuario de la Interfaz desarrollada para la predicción del Recurso Eólico disponible en la ciudad de Cúcuta	152
Anexo 7. Certificación Platzi en el curso de ANÁLISIS DE NEGOCIOS PARA CIENCIA DE DATOS	157
Anexo 8. Certificación Platzi en el curso de BÁSICO DE PYTHON	159
Anexo 9. Certificación Platzi en el curso de entorno de trabajo para CIENCIA DE DATOS CON JUPYTER NOTEBOOKS Y ANACONDA	161

1. Introducción

El presente proyecto se enfoca en la creación de un modelo que realice predicciones del Recurso Eólico promedio y total en distintos sectores de la ciudad de Cúcuta. Estas predicciones se realizaron a partir de los parámetros meteorológicos de velocidad de viento y temperatura. Para la parte del entrenamiento del modelo los datos se tomaron de la página oficial del Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM, los cuales son medidos desde la torre meteorológica del Aeropuerto Camilo Daza en la ciudad de Cúcuta, mientras que para la etapa de evaluación del modelo se utilizaron datos experimentales tomados en distintos sectores de la ciudad. En la etapa de entrenamiento se utilizó un rango de 4 años y una frecuencia de 2 minutos para los datos de velocidad del viento mientras que para los datos de temperatura se tomaron los valores promedios anuales para el mismo rango. A partir de estos datos se elaboró un modelo haciendo uso del enfoque de Aprendizaje Automático en conjunto con la técnica de Redes Neuronales Artificiales, el lenguaje Python y el entorno de programación Google Colaboratory. además, para entrenar el modelo se utilizó el algoritmo de Retropropagación con validación cruzada para evaluar los resultados del modelo, garantizando la independencia entre los datos de entrenamiento y validación. Por último, el número de épocas y la cantidad de datos con que se alimente el algoritmo son parámetros que se variaron con el fin de aumentar la precisión del modelo, la cual se analizó por medio de la métrica de éxito Error Absoluto Medio.

2. Descripción del problema

A continuación, se describe el problema a tratar en el presente documento mediante su planteamiento.

2.1. Planteamiento del problema

Es bien conocido que los combustibles fósiles como el carbón o el petróleo han sido utilizados durante un gran periodo de tiempo para generar energía. Este proceso ha traído consigo una serie de eventos que cada día son más causas de preocupación a nivel mundial en temas como: el impacto negativo al medio ambiente producido por los gases de efecto invernadero GEI, los efectos sobre la salud de las personas y la eficiencia energética. Es por esta razón que cada vez son más los países que se suman a la necesidad de fomentar el desarrollo de energías renovables (Guerrero Hoyos et al., 2020).

Las energías renovables son un tipo de energía que se genera a partir de recursos naturales como el viento, el agua y la luz solar (Giraldo et al., 2018). La energía eólica es la energía renovable que se origina por el movimiento de masa de aire que se conoce como viento. Esta fuente de energía es autosustentable, limpia y no contaminante. Además, se ha convertido en una de las opciones comercialmente más atractivas para generar energía eléctrica, y para finales de 2016 se estimó que más de 90 países tenían proyectos comerciales de energía de este tipo según datos de la World Wind Energy Association (WWEA) (Delgado, D. A. P., Herrera, M. T. J., & Almeida, 2014).

La conversión de la energía del viento a energía eléctrica se realiza por medio de un aerogenerador y la potencia que puede obtenerse es proporcional al cubo de la velocidad del viento, de ahí que la velocidad media del viento sea uno de los factores más determinantes en el momento de estudiar la viabilidad de un sistema eólico, no obstante, pueden tenerse en cuenta otros parámetros como temperatura, dirección de viento, humedad y presión (Santamarta, 2007).

La desventaja de un sistema eólico es que al trabajar con parámetros meteorológicos de comportamiento naturalmente variable y de difícil predicción, podría presentar intermitencias en su funcionamiento. Por ello es importante la ubicación de zonas donde realmente pueda aprovecharse este recurso.

Teniendo en cuenta los serios inconvenientes provocados por el uso de energías no renovables y las ventajas que ofrece el Recurso Eólico se ha llegado a la siguiente pregunta de investigación: ¿De qué manera se pueden identificar zonas estratégicas para el aprovechamiento del Recurso Eólico disponible en la ciudad de Cúcuta por medio de Ciencia De Los Datos e Inteligencia Artificial?

3. Justificación

La contaminación atmosférica es actualmente uno de los problemas ambientales más severos a nivel mundial. Dentro de los efectos negativos de esta problemática se encuentra el deterioro de la calidad del aire y la salud de las personas. La ciudad de San José de Cúcuta no ha sido una excepción al problema de contaminación ambiental, tal y como da cuenta la Corporación Autónoma Regional de la Frontera Nororiental- Corponor. De hecho, se dice que, de prolongarse esta situación, podría producirse un alto grado de toxicidad perjudicial para las personas, incluso haciéndolos vulnerables a adquirir enfermedades respiratorias o cardiovasculares producto de inhalar desechos contaminantes (González, 2009). Además, el uso de combustibles fósiles para la generación de energía eléctrica produce dióxido de carbono, el cual va directamente a la atmósfera permitiendo la acumulación de gases de efecto invernadero, que es uno de los responsables del cambio climático (González, 2009).

Teniendo en cuenta lo anterior, el presente proyecto nació de la necesidad de buscar alternativas que contribuyeron a reducir la contaminación atmosférica en la ciudad de Cúcuta, la cual siendo una ciudad pequeña suele tener efectos más concentrados en la emisión de gases contaminantes. En este orden de ideas, se consideró que el estudio de zonas que contaran con datos meteorológicos favorables para el aprovechamiento del Recurso Eólico, podía contribuir e iniciar el camino hacia la futura implementación de sistemas de generación de energía eléctrica a partir del recurso de energías renovables disponibles. Esto puede contribuir un pilar fundamental para la búsqueda a futuro de una ciudad que genere energía eléctrica principalmente con el uso de energías limpias y que no impacte negativamente al medio ambiente ni afecte la salud de las personas.

3.1. Beneficios

Los beneficios globales de analizar el recurso eólico son descritos a continuación.

3.1.1. Beneficios Ambientales

Dentro de los beneficios ambientales se encontró que el Recurso Eólico al ser una fuente limpia de energía tiene menos impacto en la contaminación atmosférica que la generada por combustibles fósiles u otras fuentes de energéticas convencionales. Otros beneficios ambientales fueron: la disminución de la contaminación del agua superficial, reducción de disturbios para la extracción de combustibles fósiles y la reducción de derrames de petróleo (Badii et al., 2016).

3.1.2. Beneficios Sociales

La realización de un proyecto en el desarrollo de energías renovables permitió avanzar en la disminución a futuro de riesgos de enfermedades respiratorias o cardiovasculares de un gran número de personas que son expuestas a los gases contaminantes de la atmósfera, los cuales son producidos en gran parte por la combustión fósil.

Debido a que las Energías Renovables requieren más mano de obra que otras fuentes de energías convencionales, es posible que haya generación de empleo, especialmente en zonas rurales donde se necesiten trabajadores locales, esto puede elevar el poder adquisitivo y la calidad de vida de los habitantes de la región (Posso et al., 2014).

3.1.3. Beneficios Institucionales

Como muy bien se sabe la Inteligencia Artificial es una tecnología tan amplia como interesante, y puede usarse para presentar soluciones innovadoras prácticamente en cualquier área. El presente proyecto trabajo en el análisis del Recurso Eólico con Inteligencia Artificial y Ciencia De Los Datos para el aprovechamiento de este tipo de energía en la ciudad. Al ser un terreno poco explorado, por lo menos en la ciudad de Cúcuta, pudo esto suponer un avance para sentar bases de futuros proyectos desarrollados por otros estudiantes de la UFPS, además de la posibilidad de ser publicado en una revista de investigación.

3.1.4. Beneficios Económicos

El uso de un recurso renovable puede responder de una forma más eficiente y limpia a la creciente demanda energética producida por el aumento de la población, a la futura expansión industrial en la ciudad y el acelerado crecimiento del sector tecnológico (Londoño, 2017). La utilización eficiente de energía por la población y el Estado apoya a la economía local y disminuye las dependencias de fuentes extranjeras de energías derivadas de combustión fósil. Además, la conservación de la energía permitida por este recurso protege la economía y los consumidores de fluctuaciones en el precio e interrupciones en el servicio por causa de desastres naturales y otros tipos.

4. Alcance

La investigación que se llevó a cabo en el presente proyecto es de tipo correlacional, ya que se evaluaron dos variables meteorológicas para analizar su efecto sobre el Recurso Eólico. Debido a que dicho recurso puede depender de muchas variables meteorológicas, también es cierto que unas aportan más que otras. En este sentido, se seleccionaron las dos variables que más aportan a la estimación del Recurso Eólico; las cuales fueron velocidad del viento y temperatura. Con estas variables se construyó un modelo de Redes Neuronales Artificiales que realiza predicciones del Recurso Eólico disponible a largo plazo en la zona del Aeropuerto Camilo Daza en la ciudad de Cúcuta, el cual fue alimentado con una base de datos del Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM. Por otra parte, con el modelo ya funcional se realizaron predicciones a corto plazo sobre otras zonas de la ciudad a partir de la toma de datos experimentales. Además, se realizó una interfaz donde se puede observar el comportamiento de los datos y sus predicciones. Dentro de los logros esperados de este proyecto se encuentra el hallar una técnica innovadora en el análisis del Recurso Eólico a nivel local.

5. Limitaciones y delimitaciones

Las limitaciones y delimitaciones del proyecto son explicadas en esta sección.

5.1. Limitaciones

El presente proyecto realizó un análisis del Recurso Eólico en la ciudad de Cúcuta mediante Inteligencia Artificial, haciendo uso específicamente de la rama Ciencia de los Datos y Aprendizaje Automático, donde se entrenó al sistema con una base de datos. Antes de especificar las herramientas que se utilizaron es importante resaltar que el uso de softwares y equipos de medición de los laboratorios de la UFPS se encontraban aún inhabilitados por el cierre del campus debido al confinamiento. En este sentido se seleccionaron y adquirieron por cuenta propia las herramientas de softwares, equipos de medición y recursos en general necesarios para el desarrollo del proyecto tanto a nivel de cómputo como a nivel experimental. Dentro de las herramientas de software seleccionadas para trabajar Ciencia de los datos, desde el preprocesamiento, análisis y creación del modelo predictivo, se utilizaron los entornos en línea de Google Colab y Jupyter Lab, mientras que para el desarrollo de la Interfaz de usuario se utilizó el IDE de escritorio PyCharm. Por otra parte, para la realización de pruebas de funcionamiento sobre la medición de velocidad de viento y temperatura para la estimación del Recurso Eólico disponible se utilizó un sensor digital conocido como anemómetro. Todas las herramientas computacionales y técnicas seleccionadas fueron de software libre.

5.2. Delimitaciones

Las delimitaciones globales del proyecto son descritas a continuación.

5.2.1. Espacio geográfico y tiempo

Este proyecto se realizó en la ciudad de Cúcuta. Cabe destacar que se seleccionaron las zonas de Torcoroma, Ceiba y Belén las cuales se realizaron las mediciones de velocidad de viento y

temperatura para la estimación del Recurso Eólico disponible. Con respecto a la duración del proyecto se delimitó para un período de 6 meses.

5.2.2. Variables

Aunque las variables meteorológicas son más importantes en conjunto para determinar con precisión una medida en específico, sí que es cierto que existen unas variables que aportan más que otras. En este sentido se eligieron las dos variables meteorológicas que más aportan a la estimación del Recurso Eólico disponible, las cuales fueron velocidad de viento y temperatura.

5.2.3. Profundidad del tema

El desarrollo de este proyecto se basó en la disciplina de Inteligencia Artificial y específicamente en las ramas de Ciencia de los Datos y Aprendizaje Automático. Se manejó una base de datos para el entrenamiento del modelo predictivo mediante Redes Neuronales Artificiales con herramientas de Aprendizaje Automático en Python, donde el programa aprendió a reconocer patrones y a realizar predicciones a corto y a largo plazo.

6. Objetivos

6.1. General

Analizar el Recurso Eólico en la ciudad de Cúcuta mediante Ciencia De Los Datos e Inteligencia Artificial.

6.2. Específicos

- Recopilar información sobre Recurso Eólico e Inteligencia Artificial.
- Seleccionar la tecnología y la herramienta de software para la Inteligencia Artificial.
- Crear el modelo de Ciencia de los Datos e Inteligencia Artificial
- Evaluar los resultados para determinar las zonas estratégicas donde se pueda aprovechar el Recurso Eólico en la ciudad de Cúcuta.

7. Marco referencial

Esta sección del documento está orientada a cubrir los aspectos y enfoques más importantes para el entendimiento de esta propuesta, así como exponer proyectos similares tanto a nivel internacional, nacional y local.

7.1. Antecedentes

En lo que concierne a la temática abordada en el presente proyecto de investigación, diferentes teóricos dieron a conocer sus puntos de vista y aportes, así mismo se encontraron investigaciones que profundizan acerca del tema a nivel internacional, nacional y local ya que es un tema de gran interés científico.

Dentro de los proyectos realizados a nivel internacional, se encuentra la temática inteligencia artificial que lleva por título Algoritmos De Aprendizaje Automático Para Análisis Y Predicción De Datos, realizada por Lilian Judith Sandoval en el Salvador durante el año 2018. En esta investigación se utiliza la información obtenida digitalmente para realizar un análisis de comportamiento y predecir tendencias futuras. Este aporte resultó fundamental para observar el funcionamiento de un algoritmo de Aprendizaje Automático, lo cual fue de suma importancia para el presente proyecto donde se hizo uso de herramientas de Aprendizaje Automático y Ciencia De Los Datos en la anticipación del comportamiento de variables relacionadas con el Recurso Eólico (Sandoval Serrano, 2018).

Ahora bien, una investigación realizada en México para el año 2018 denominada “Predicción de potencia eólica utilizando técnicas modernas de Inteligencia Artificial” hizo un pronóstico del potencial eólico mediante Redes Bayesianas; la cual es una técnica de Inteligencia Artificial basado en un modelo probabilístico, y que permite manejar una serie de variables aleatorias, así como conocer la relación entre las mismas. La investigación también resalta métodos de predicción

más allá de Inteligencia Artificial como series de tiempo y métodos estadísticos. De esta forma se pudo observar el uso específico de una técnica de Inteligencia Artificial, así como información adicional sobre otros métodos; esto fue de mucho interés para comprender que los métodos pueden complementarse para realizar comprobaciones. Sin embargo, el aporte principal de esta investigación radica en que los modelos de Inteligencia Artificial suelen manejar una mejor precisión para predecir velocidades de viento, el cual fue un factor determinante en la elección de este proyecto, ya que este parámetro es fundamental para la predicción del Recurso Eólico (Predicción de Potencia Eólica Utilizando Técnicas Modernas de Inteligencia Artificial, n.d.).

Así mismo, en México una investigación llamada “Técnicas de aprendizaje automático en el diagnóstico de aerogeneradores” realizada en el año 2019, describe como El Centro Mexicano de Innovación en Energía Eólica (CEMIE-Eólico) realizó un sistema de diagnóstico de aerogeneradores con la utilización del método SCADA. Para entrenar a este sistema de aprendizaje automático se tomaron los datos de varias simulaciones que presentaban fallas en el aerogenerador, esto permite al sistema familiarizarse por medio de datos de validación. Mas allá del tipo de técnica utilizada para el Aprendizaje Automático, este tipo de investigación fue de ayuda para resaltar la importancia de considerar los datos de validación como una parte importante en el proceso de entrenamiento de un modelo que realice predicciones (García et al., 2019).

En Quito, Ecuador, un artículo de investigación denominado “Plataforma Web de predicción de viento basado en técnicas de inteligencia artificial con el uso del método ARIMA: Caso de estudio” y publicado por la Revista Ibérica de Sistemas y Tecnologías de Información, expone el desarrollo de una aplicación que predice el análisis del viento. Esta plataforma de predicción trabaja con un conjunto de datos obtenidos de una torre meteorológica ubicada en la cordillera de Los Andes en Ecuador. La torre proporciona información de temperatura, presión atmosférica,

dirección de viento, humedad relativa, radiación solar y otras variables asociadas a la predicción del viento. Para el análisis y la predicción de los datos se utilizó el método ARIMA y lenguaje R que es un entorno enfocado al análisis estadístico. El aporte de esta investigación tuvo su importancia en reconocer otros métodos para la predicción de datos con inteligencia artificial, así como todo el proceso que involucró la toma de los datos con los cuales el sistema realizó las predicciones, las cuales en ese caso fue tomada de la torre meteorológica del IDEAM en el Aeropuerto Camilo Daza (Ramírez et al., 2019).

El artículo de investigación llamado “Redes neuronales en la predicción de la velocidad del viento” realizado en 2020 en Ecuador, desarrolla un sistema de predicción del viento mediante el uso de Redes Neuronales para estimar la cantidad de energía eólica que se dispone. El sistema utilizó los datos de la estación meteorológica de Chimborazo, propiedad de ESPOCH para el entrenamiento de la Red Neuronal. Una vez realizado el aprendizaje automático el sistema es capaz de predecir en dos formas: con datos conocidos por la red y con datos desconocidos por la red (nunca utilizados en el entrenamiento). Aunque los softwares y/o modelos utilizados en este artículo pudieron variar con respecto a los del presente proyecto, existieron aportes muy amplios como la descripción del proceso de entrenamiento de la Red Neuronal con datos reales, lo cual fue de mucha ayuda para comprender los factores relacionados con el aprendizaje automático del modelo creado propiamente (I, 2020).

A nivel nacional, la investigación llamada “Estudio estadístico de la velocidad y la dirección del viento en los departamentos de Atlántico y Bolívar en Colombia” realizado en el año 2018, el cual resalta que la generación de energía eléctrica en Colombia a partir de energías renovables es muy poca, siendo la eólica con la mayor participación con un 0.1% de la demanda eléctrica. El estudio analiza las variaciones de variables eólicas como velocidad y dirección de viento para

estimar si el potencial suministrado por el recurso eólico puede ser suficiente para abastecer a la región a partir de parques eólicos. Este estudio tuvo un alto aporte para el presente proyecto ya que ambos están en la dirección de estudiar el aprovechamiento del Recurso Eólico disponible en una región específica en Colombia, aunque utilizando métodos distintos (de la Cruz Buelvas et al., 2018).

7.2. Marco Teórico

A continuación, se expondrán las generalidades más relevantes de este proyecto.

7.2.1. Recurso Eólico

El recurso eólico aprovecha la energía producida por el movimiento del viento y esta transformando ésta en energía mecánica o eléctrica. Durante la historia el hombre ha utilizado el recurso eólico para su beneficio, por ejemplo para el desplazamiento de los barcos de vela o para el bombeo de agua. Para convertir la energía del recurso eólico en energía eléctrica se utilizan los aerogeneradores, los cuales existen en gran variedad de tipos según la orientación de sus palas; el más utilizado es el de eje horizontal (Fe, 2010).

El principal factor para la generación de energía eólica es el viento, el cual es generado por el movimiento de la masa atmosférica. Sin embargo para la estimación del potencial eólico es necesario conocer no sólo la velocidad del viento sino también otros parámetros como la dirección del viento, densidad y humedad atmosférica (Camacho, n.d.)(Encinas, E. Torrez, R. Ghezzi, F. Palenque, 2015).

La forma en que se suele estimar el Recurso Eólico es mediante la potencia que se puede extraer del viento. Si se tiene una masa de aire m con velocidad v , la energía cinética generada estará dada por la ecuación (1).

$$E = \frac{1}{2}mv^3 \quad (1)$$

El caudal de masa de aire m de densidad ρ que fluye con velocidad v por medio de una superficie de área A ortogonal a la dirección del flujo se puede observar en la ecuación (2).

$$Pd = \frac{1}{2}mv^3 = \frac{1}{2}\rho Av^3 \quad (2)$$

La densidad de potencia disponible o potencia disponible por unidad de área (Pd/A) se puede apreciar en la ecuación (3).

$$\frac{Pd}{A} = \frac{1}{2}\rho v^3 \quad (3)$$

Donde Pd es la potencia disponible, A es el área, ρ es la densidad del aire seco en y v es la velocidad del viento en $[m/s]$. Para calcular la densidad del aire seco se utiliza la siguiente ecuación (4).

$$\rho = \frac{Patm}{Rp.Tk} \quad (4)$$

Donde $Patm$ es la presión atmosférica, Rp es la constante universal sobre masa molecular del gas para el aire y Tk es la temperatura en Kelvin. Por otro lado, la presión atmosférica $Patm$ se calcula mediante la ecuación (5).

$$Patm = Po.e^{-\left(\frac{g.Z}{Rp.Tk}\right)} \quad (5)$$

Donde Po es la presión a nivel del mar es $[Pa]$, g es la gravedad en $[m/s^2]$, Z es la altura sobre el nivel del mar de la zona en $[m]$. Por ultimo la constante Rp y la variable Tk definida anteriormente como la constante universal sobre masa molecular del gas para el aire y la temperatura en Kelvin respectivamente.

Si se sustituye la ecuación (5) en (4) y en (3) se obtiene la ecuación (6).

$$\frac{Pd}{A} = \frac{Po \cdot e^{-\left(\frac{g \cdot Z}{Rp \cdot Tk}\right)} \cdot v^3}{2 \cdot Rp \cdot Tk} \left[\frac{W}{m^2} \right] \quad (6)$$

7.2.2. Variación de la velocidad media a largo plazo con la altura

Este caso corresponde a la velocidad media a largo plazo como velocidad media anual, mensual, etc. El modelo más utilizado es basado en la ley de Hellman para la variación de la velocidad con la altura, según la ecuación (7).

$$v' = v \left(\frac{z'}{z} \right)^\alpha \quad (7)$$

Donde α es un coeficiente que depende de la longitud de rugosidad (Z_0) del terreno, v es la velocidad conocida a la altura z , mientras que z' es la altura a la cual se quiere hallar la velocidad v' (Miguel Villarrubia López, 2012).

7.2.3. Variación de la velocidad del viento estacionario con la altura

La velocidad del viento medio o estacionario se obtiene a partir del promedio de velocidad para intervalos de tiempo de 10 minutos (Miguel Villarrubia López, 2012). Su variación con respecto a la altura se corresponde a la ecuación (8).

$$v' = v \frac{\ln\left(\frac{z'}{Z_0}\right)}{\ln\left(\frac{z}{Z_0}\right)} \quad (8)$$

En donde v , v' , son respectivamente las velocidades a las alturas z y z' , en (m/s). Z_0 es la rugosidad del terreno en (m).

7.2.4. Inteligencia Artificial y Aprendizaje Automático

La Inteligencia Artificial puede definirse de muchas maneras ya que se considera hoy en día un campo multidisciplinario. Una de sus definiciones la cataloga como la capacidad de las máquinas para simular el cerebro humano, por medio del proceso de aprender de los datos para la toma de

decisiones. La ventaja principal de la Inteligencia Artificial frente a los seres humanos es que la máquina no necesita descanso y, además, puede analizar grandes volúmenes de datos al mismo tiempo, lo cual para un ser humano resultaría sumamente difícil, por no decir imposible. Además, las máquinas tienen un índice de error mucho menor al que puede tener un ser humano. A pesar de que para muchos el concepto de Inteligencia Artificial puede sonar futurista, lo cierto es que esta disciplina ya está prácticamente en todos los sectores de la tecnología, como por ejemplo en el reconocimiento facial en temas de seguridad, las recomendaciones que hacen las redes sociales en función de cómo utilizamos las aplicaciones, las predicciones del tiempo, detección y clasificación de objetos, protección contra amenazas cibernéticas, entre muchísimas más.

Dentro de los enfoques más importantes de la Inteligencia Artificial se encuentra el Aprendizaje Automático, el cual está relacionado con la capacidad de aprendizaje sin estar programados para ello. Ejemplos de algunas aplicaciones del Aprendizaje Automático son los resultados del motor de búsqueda de Google y los filtros de *spam* de correo electrónico. Dentro de los mecanismos de Aprendizaje Automático se encuentran:

- Aprendizaje supervisado: Usan datos que ya han sido organizados con anterioridad con el fin de establecer un patrón para categorizar los datos nuevos. Se requiere la intervención humana.
- Aprendizaje no supervisado: No usan datos clasificados previamente y el algoritmo tiene que buscar la forma de categorizar la nueva información por sí mismo. No requiere intervención humana.
- Aprendizaje por refuerzo: Este aprendizaje se basa en la experiencia. Es necesario darle un refuerzo positivo al algoritmo siempre que acierte, ver Fig. 1. Tipos de Aprendizaje Automático..



Fig. 1. Tipos de Aprendizaje Automático. (Rouhiainen, n.d.).

7.2.5. Aprendizaje Profundo

El Aprendizaje Profundo es una de las herramientas más poderosas de la Inteligencia Artificial y es a la vez un subcampo del Aprendizaje Automático. EL Aprendizaje Profundo busca analizar y resolver problemas de gran dificultad con un alto volumen de datos. Esta disciplina se usa comúnmente con redes neuronales, las cuales se componen de capas y que a su vez reconocen patrones complejos en los datos (Rouhiainen, n.d.).

7.2.6. Ciencia de los Datos

La Ciencia de los Datos actualmente es un recurso elemental en el análisis y explotación de los datos. La finalidad de esta herramienta se encuentra en la búsqueda de modelos de reconocimiento de patrones y comportamientos a partir de los datos con el propósito de tomar sus propias decisiones y realizar predicciones. Es un sector que ha evolucionado al tener en cuenta el acceso a grandes cantidades de datos e incluso su manejo en tiempo real, por medio de técnicas complejas que pueden tratar con los problemas como escalabilidad, robustez ante errores, adaptabilidad con modelos dinámicos, etc (Encinas, E. Torrez, R. Ghezzi, F. Palenque, 2015).

En los últimos años la Ciencia de los Datos surgió como una herramienta moderna e importante.

Podría ser visto como la unión de varias disciplinas clásicas como estadística, minería de datos, bases de datos y sistemas distribuidos. Las áreas existentes deben unirse para convertir los datos disponibles en valor para las personas, organizaciones y la sociedad (Fernández Estrella et al., 2015).

7.2.7. Redes Neuronales Artificiales

Uno de los sectores más destacados del ámbito científico de la Inteligencia Artificial es el correspondiente a las Redes Neuronales Artificiales (RNAs). Las RNAs tratan de imitar la conducta del cerebro humano, por medio de la experiencia y la adquisición de conocimiento a través de un conjunto de datos. También pueden considerarse modelos computacionales que han surgido como un esfuerzo por formular matemáticamente la estructura del cerebro tratando de evitar los fallos o errores humanos. Los componentes básicos de la estructura de una RNA son las neuronas y se agrupan por medio de capas. Estas capas agrupan un grupo de neuronas provenientes de la misma entrada y cuyas salidas tienen el mismo destino.

Las neuronas artificiales también llamadas procesadores elementales ponderan las entradas que recibe, es decir le asigna ciertos pesos, los cuales son fundamentales para el aprendizaje de la red. De esta manera, la RNA puede aprender incluso de sus propios errores por medio de un procedimiento inductivo fundamentado en la muestra de un grupo de patrones informativos que dan lugar a que el sistema extienda generalice conceptos a través de casos particulares.

Según el tipo de arquitectura, una red neuronal artificial puede estar compuesta por varias capas, de tal forma que cada capa está conectada a la capa siguiente, a las que pueden transmitir información, ver Fig. 2. Estructura jerárquica de un sistema basado en RNA.. Estas capas se clasifican es:

- Capas de entrada: También denominadas sensoriales, son las encargadas de recibir datos o

señales del exterior. Su equivalente en una red neuronal biológica son las dendritas.

- Capas ocultas: Se encargan de llevar a cada todo el trabajo de la red. Estas capas no tienen conexión con el entorno. En esta capa se le proporciona la libertad necesaria a la red para que sea capaz de representar los aspectos del entorno que desea modelar.
- Capas de salida: Estas capas se componen de una cantidad de neuronas las cuales suministran la respuesta de la red (Larranaga, n.d.-b; Miranda, n.d.)

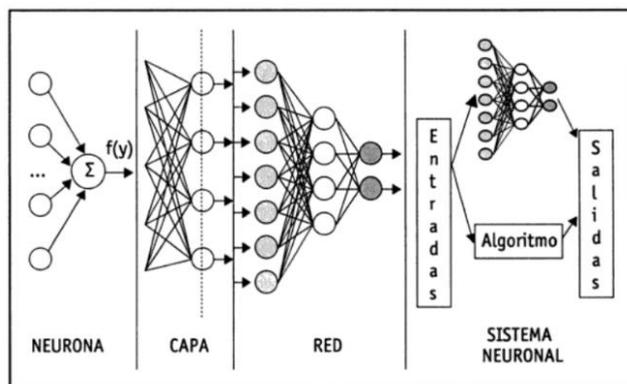


Fig. 2. Estructura jerárquica de un sistema basado en RNA. (Zhang, 2020).

7.2.8. Perceptrón simple

El perceptrón simple es un modelo de red neuronal que consta de dos capas, una de entrada y una de salida. Las neuronas de entrada son discretas y la función de activación es de tipo escalón. Este perceptrón, debido a la naturaleza de su funcionamiento, suele utilizarse como clasificador. En la Fig. 3. Arquitectura (izquierda) y función de transferencia (derecha) de un perceptrón simple.. Se puede observar la arquitectura y la función de transferencia de un perceptrón simple.

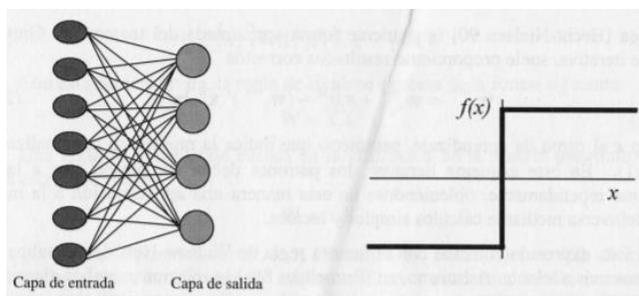


Fig. 3. Arquitectura (izquierda) y función de transferencia (derecha) de un perceptrón simple. (Larranaga, n.d.-b).

El perceptrón simple está limitado por el hecho de que sólo puede representar funciones linealmente separables, ver Fig. 4. Región de decisión correspondiente a un perceptrón simple..

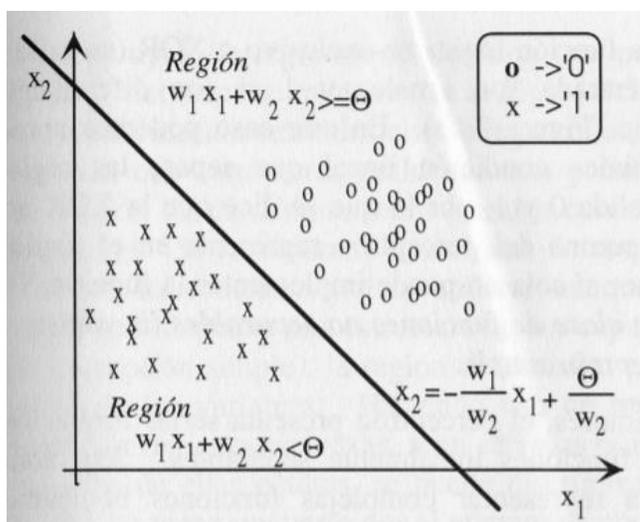


Fig. 4. Región de decisión correspondiente a un perceptrón simple. (Larranaga, n.d.-b).

7.2.9. Perceptrón multicapa

En el perceptrón multicapa se incluyen las capas ocultas en la red permitiendo que la región de decisión pueda discriminar regiones convexas, a diferencia del perceptrón simple en donde la región de decisión es una línea recta. Este perceptrón multicapa o MLP (Multi-Layer Perceptron) suele ser entrenado a través de un algoritmo de retropropagación de errores o BP (Back Propagation), es por ellos que la arquitectura de este perceptrón es también conocida como red de retropropagación. Fig. 5 Regiones de decisión obtenidas para el perceptrón simple (arriba), el perceptrón multicapa con una capa oculta (en medio) y el perceptrón multicapa con dos capas

ocultas (abajo).. Presenta las regiones de decisión para distintas arquitecturas con dos neuronas en la capa inicial.

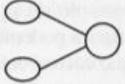
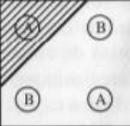
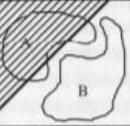
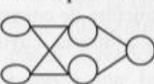
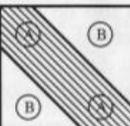
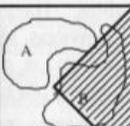
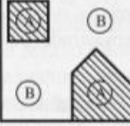
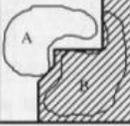
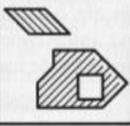
Arquitectura	Región de decisión	Ejemplo 1: XOR	Ejemplo 2: clasificación	Regiones más generales
Sin capa oculta 	Hiperplano (dos regiones)			
Una capa oculta 	Regiones polinomiales convexas			
Dos capas ocultas 	Regiones arbitrarias			

Fig. 5 Regiones de decisión obtenidas para el perceptrón simple (arriba), el perceptrón multicapa con una capa oculta (en medio) y el perceptrón multicapa con dos capas ocultas (abajo). (Larranaga, n.d.-a).

La arquitectura y la función de activación del MLP con una capa oculta se puede observar en la Fig. 6. Arquitectura (izquierda) y función de activación (derecha) para el perceptrón multicapa..

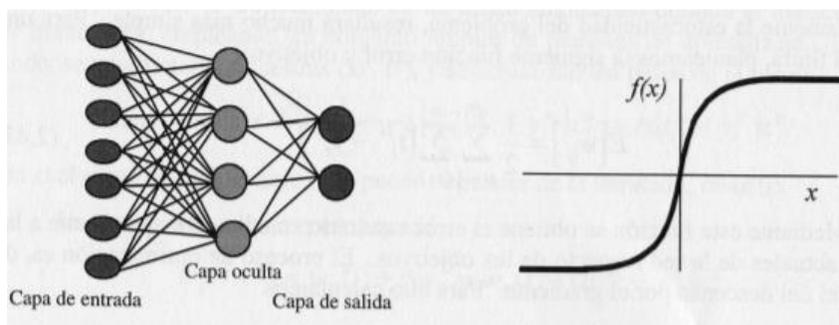


Fig. 6. Arquitectura (izquierda) y función de activación (derecha) para el perceptrón multicapa. (Larranaga, n.d.-a).

7.2.10. Retropropagación

Esta red lleva este nombre debido a que el error es propagado hacia atrás desde la capa de salida, es por ello que los pesos de las conexiones de las neuronas de la capa oculta cambian en el durante el entrenamiento. Este cambio de pesos en las neuronas puede influir en la entrada global, la

función de activación y la salida, por esta razón es importante tener en cuenta las variaciones que presentan los resultados al variar el valor de los pesos. A esto se le conoce como sensibilidad de la función de activación (Alberto Ruiz Marta Susana Basualdo Autor & Jorge Matich, n.d.).

7.2.11. Validación K-Fold

Es un método que permite dividir aleatoriamente los datos de entrenamiento en k particiones, en donde una parte se utilizará para el entrenamiento y otra parte para la validación. Este proceso se realizará durante k iteraciones, de forma que cada partición será utilizada para el entrenamiento (Universitat de Valencia, 2018).

7.2.12. Python

Python es un lenguaje de programación de alto nivel multi-paradigmas y de software libre que maneja una gran cantidad de herramientas que hacen que su uso sea simple y a la vez completo. En este lenguaje se puede trabajar la programación orientada a objetos, así como funcional e imperativa (Larranaga, n.d.).

Python además reúne las herramientas necesarias para trabajar con Ciencia De Los Datos (Data Science) y un conjunto de librerías para uso matemático, Machine Learning y Ciencia en general (Miranda, n.d.).

7.2.13. Anemómetro de álabes

El anemómetro es un instrumento de medición que permite medir la rapidez del viento en un instante determinado. Estos sensores suelen incluir también la medición de otros parámetros atmosféricos como: temperatura, humedad, altitud y presión atmosférica. El anemómetro de álabes consiste en medidor de velocidad mecánica rotatorios y tiene una forma de álabes o hélice. Algunos de estos sensores, como es el caso del seleccionado para el proyecto, toman las lecturas de la velocidad del aire y de temperatura de forma discreta, captando los impulsos eléctricos

suministrados por un transductor inductivo o capacitivo. El eje de rotación suele ser horizontal y paralelo a la dirección del viento, ver Fig. 7. Anemómetro de álabes.. Este anemómetro fue seleccionado por las siguientes razones:

- Fácil uso.
- Costo accesible.
- Toma de mediciones de velocidad del aire y temperatura.
- Permite exportar los datos y visualizar las gráficas del comportamiento de las variables en los periodos de medición (Faustino Menéndez Díez, 2006).



Fig. 7. Anemómetro de álabes. (Faustino Menéndez Díez, 2006).

7.3. Marco Legal

El presente trabajo de investigación toma como base la normatividad a nivel nacional e internacional lo siguiente:

- **Ley 143 de 1994 Diario Oficial No. 41.434, del 12 de julio de 1994.**

Por la cual se establece el régimen para la generación, interconexión, transmisión, distribución

y comercialización de electricidad en el territorio nacional, se conceden unas autorizaciones y se dictan otras disposiciones en materia energética.

Artículo 7°. Podrán formar parte de las actividades del sector diversos agentes económicos, públicos, privados o una combinación de ellos para llevar a cabo funciones en un entorno de libre competencia, en conformidad con los artículos 333, 334 el penúltimo inciso del artículo 336 de la Constitución Nacional, y el artículo 3°. De esta Ley. En los casos especificados por la ley, para desarrollar los proyectos se deberán adquirir los respectivos permisos de las autoridades competentes en temas ambientales, sanitarios, usos de aguas y los de carácter municipal que sean exigibles.

En las actividades del sector podrán participar diferentes agentes económicos, públicos, privados o mixtos, los cuales gozarán de libertad para desarrollar sus funciones en un contexto de libre competencia, de conformidad con los artículos 333, 334 y el inciso penúltimo del artículo 336 de la Constitución Nacional, y el artículo 3° de esta Ley. En los casos señalados por la ley, para operar o poner en funcionamiento los proyectos se deberán obtener de las autoridades competentes los permisos respectivos en materia ambiental, sanitaria, uso de aguas y los de orden municipal que sean exigibles (*Ley_143_1994 (1)*, n.d.).

- **Ley de 2002 normatividad para las Energías Renovables en Colombia**

Esta ley permite la prerrogativa sobre los impuestos de renta para ingresos generados a partir de comercialización de la red eléctrica derivada de residuos agrícolas, recurso eólico y biomasa. Se debe cumplir con los requerimientos de tramitar certificados de emisión de CO₂ y que un mínimo de 50% de los recursos generados por el comercio sean destinados a la inversión de proyectos en beneficio a la sociedad (Cámara de Comercio de Cali, 2016).

- **Decreto 3683 de 2003 nivel nacional por el cual se reglamenta la Ley 697 de 2001 y se**

crea una Comisión Intersectorial

Este decreto regula el uso y la eficiencia de la energía incluyendo las fuentes no convencionales que son ambientalmente autosustentables como la energía solar, energía eólica, energía geométrica, entre otros. También crea estímulos para la investigación sobre el desarrollo de estas energías y fija méritos a los que hagan un eficiente y racional de la misma. Por otro lado, también resalta lineamientos para las empresas de servicios públicos, así como lo que debe expresarse en los recibos del servicio eléctrico, gas y los derechos de los consumidores (Secretaría Jurídica Distrital, 2003).

- **Ley 1715 de 2014 aplicación de los incentivos tributarios Ministerio de Minas y Energías**

La finalidad de esta ley radica en promover el uso de las fuentes de energías renovables en el país y de esta forma, aumentar la seguridad al abastecer la demanda energética, así como disminuir la emisión de gas invernadero. El Ministerio de Minas y Energías, por medio de la Unidad de Planeación Minero-Energética (UPME), el ministerio de Ambiente y Desarrollo Sostenible, la Autoridad Nacional de Licencias Ambientales (ANLA) y la Agencia de Estados Unidos para el Desarrollo Internacional (USAID) por medio del programa de Energía Limpia para Colombia (CCEP), han trabajado en una convocatoria con el propósito de que empresarios, inversionistas y ciudadanos se unan a la apuesta por el uso de energías renovables. Cabe destacar que para los que deseen involucrarse en esta iniciativa deberán cumplir con una serie de requisitos especificados por la entidad pertinente (*Invierta y Gane Con Energía Guía Práctica Para La Aplicación de Los Incentivos Tributarios de La Ley 1715 de 2014*, n.d.).

8. Diseño metodológico

El diseño metodológico se realizó con respecto a las actividades que tienen como finalidad el cumplimiento de cada uno de los objetivos específicos.

8.1. Objetivo 1. Recopilar información sobre Recurso Eólico e Inteligencia Artificial

El primer objetivo estuvo relacionado con la familiarización de los conceptos sobre los que el proyecto sentó sus bases. Para ello se llevó a cabo una investigación sobre el Recurso Eólico enfocado principalmente en la ciudad de Cúcuta. Además, se consultó si existen bases de datos sobre variables relacionadas con el Recurso Eólico en la ciudad con el fin de obtener los valores que utilizará el sistema. Por otro lado, fue necesario realizar una investigación general sobre Inteligencia Artificial con el fin de relacionarnos con los conceptos y técnicas que dieron solución a la problemática planteada. Por último, se seleccionó un gestor de referencias para guardar y ordenar toda la información recopilada, y de esta forma tener acceso a ella cuando sea necesario.

Actividades:

- Recopilación de información sobre recurso eólico en la ciudad de Cúcuta.
- Consulta de bases de datos relacionado con el recurso eólico en la ciudad de Cúcuta.
- Investigación general sobre inteligencia artificial.
- Selección de un gestor de referencias.

8.2. Objetivo 2. Seleccionar la tecnología y herramienta de software para la Inteligencia Artificial

Para el cumplimiento del segundo objetivo es importante conocer que tipos de tecnologías maneja la Inteligencia Artificial, donde fue necesario revisar conceptos como Aprendizaje Automático y Ciencia De Los Datos. La investigación debe extenderse hasta contar con una serie de referencias sobre sistemas que trabajen con datos de Recurso Eólico haciendo uso de la

Inteligencia Artificial. Toda esta información se procederá a clasificarse para realizar una comparación sobre las distintas herramientas, tecnologías, softwares y conceptos con el fin de definir cuáles son las que mejor se adaptan a las necesidades del proyecto. Para la selección de las herramientas de software también se tuvo en cuenta la verificación de tipo de licencia, compatibilidad con los computadores con que se cuenten, y el nivel de dificultad para su manejo.

Actividades:

- Recopilación sobre las distintas tecnologías y sistemas eólicos con IA.
- Comparación de las distintas tecnologías y softwares investigados para definir cuál se adapta mejor a las necesidades del proyecto.

8.3. Objetivo 3. Creación del modelo de Ciencia de los Datos e Inteligencia Artificial

En este objetivo se desarrolló un modelo capaz de realizar predicciones de Recurso Eólico con el fin de analizar el comportamiento a corto plazo del mismo en distintas zonas de la ciudad. Para el cumplimiento de este objetivo fue necesario primero que nada descargar la base de datos de los parámetros seleccionados para estimar el Recurso Eólico, en este caso fueron velocidad de viento y temperatura. Posteriormente en el preprocesamiento de los datos se aseguró que los mismos estén en el mismo formato para alimentar el algoritmo. Además, en la exploración de los datos se revisó si existían datos faltantes, valores atípicos y reconocimiento de patrones que faciliten la creación del modelo. Luego el algoritmo de aprendizaje del modelo se entrenó con los datos ya limpios y procesados en las etapas anteriores. La evaluación del algoritmo se hizo mediante métricas de éxito como el Error Absoluto Medio del Inglés MAE para estimar la precisión de las predicciones. Una vez que las predicciones fueron lo suficientemente precisas se dio paso a usar el modelo. Para darle utilidad al algoritmo se utilizaron datos experimentales para realizar las respectivas pruebas de funcionamiento. Además, se creó una interfaz de usuario para visualizar los resultados tanto

teóricos como experimentales.

Actividades:

- Colección de datos
- Preprocesamiento de datos
- Exploración de datos
- Entrenamiento del algoritmo
- Evaluación del algoritmo
- Uso del modelo

8.4. Objetivo 4. Evaluar los resultados para determinar las zonas estratégicas donde se pueda aprovechar el recurso eólico en la ciudad de Cúcuta

Para dar cumplimiento al objetivo fue necesario realizar un análisis de resultados tanto a la interfaz como a los datos obtenidos experimentalmente. En caso de presentar algún error se deberá buscar solución inmediatamente. Luego, la comparación de estos resultados será de utilidad para verificar la precisión con la que trabaja la interfaz en comparativa a los valores tomados experimentalmente. Una vez se conozca la fiabilidad de la interfaz, se identificarán que zonas cumplen con las condiciones para el aprovechamiento del recurso eólico.

Actividades:

- Análisis de resultados obtenidos.
- Realizar correcciones si es necesario.
- Identificación de las zonas donde se cumplan las condiciones para el aprovechamiento del recurso eólico.

9. Resultados

En esta sección se presentan los resultados obtenidos de acuerdo con los objetivos específicos.

9.1. Estado del arte sobre Recurso Eólico e Inteligencia Artificial

Se consultaron diversas fuentes a fin de comprender el aprovechamiento del recurso eólico y el uso de la inteligencia artificial para el tratamiento de datos y el entrenamiento de algoritmos de predicción. Entre estos se encuentra “Aprendizaje automático en la identificación de sistemas. Un caso de estudio en la predicción de la generación eléctrica de un parque eólico” el describe la implementación de un sistema predictivo a partir de algoritmos de aprendizaje supervisado, para estimar la generación de potencial eléctrico en un campo eólico, todo esto usando el lenguaje de programación Python (Aguilar et al., 2019).

Por otro lado, se encuentra “Evaluación y Caracterización Del Recurso Eólico en la Universidad Francisco de Paula Santander, Cúcuta, y Prospectiva para el aprovechamiento energético en el Norte de Santander” el cual destaca el análisis de la velocidad y dirección del viento, temperatura y presión utilizando la distribución de Weibull en Matlab y así comprender el comportamiento de la velocidad del viento en Cúcuta mediante una serie graficas (Autónoma de Occidente Colombia Vera Duarte et al., n.d.).

Determinar la viabilidad de la producción de energía eólica es algo esencial ya que podremos establecer si es factible la implementación de aerogeneradores, por esta razón se consultaron diferentes metodologías, entre estas se encuentra “Metodología para determinar la viabilidad de generación de energía eléctrica promedio del recurso eólico” en el que se realiza un análisis estadístico de una extensa base de datos para determinar si en dicha zona es óptimo implementar una estación eólica de baja potencia (Henaon león et al., 2018).

Existen diferentes modelos que son usados para predicción de velocidad del viento, cada

modelo posee sus respectivas ventajas y desventajas, por lo cual se consultó “Revisión de la literatura de modelos computacionales para la predicción de la velocidad del viento de 2004 a 2016” el cual reúne los diversos modelos que son utilizados en la actualidad y enfocados a diferentes escenarios, además, detalla las ventajas y desventajas de cada uno de estos en sus respectivos lenguajes de programación (Tripp-Barba, 2020).

9.1.1. Bases de datos para el Recurso Eólico

La base de datos utilizada se obtuvo del Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM el cual es dependiente del Ministerio de Ambiente y Desarrollo Sostenible del gobierno de Colombia. Los datos extraídos son medidos específicamente en la torre meteorológica del Aeropuerto Camilo Daza en la ciudad San José de Cúcuta (IDEAM, n.d.).

9.1.2. Selección del gestor de referencia

Para realizar esta selección se reunió información de los gestores de referencia más utilizados (*Deusto. Biblioteca Universitaria. Biblioguías LibGidak*, n.d.) y se compararon mediante una matriz de Pugh. La selección se realiza mediante la comparación de parámetros a los que se les asigna un peso específico y seguidamente se comparan entre sí, la asignación del peso será de la siguiente manera: se le asigna un puntaje de 10 si la opción es mucho más eficiente que la comparada, se le asigna un puntaje de 5 si es más eficiente que la comparada, se le asigna un puntaje de 1 si es igual a la comparada, se le asigna un puntaje 0.2 si es menos eficiente que la comparada y se le asigna un puntaje de 0.1 si es mucho menos eficiente que la comparada (Vidal et al., n.d.). En la Tabla 1. Asignación de puntaje. se puede observar la asignación del puntaje.

Tabla 1. Asignación de puntaje.

ESCALA VERBAL	ESCALA NUMÉRICA
Mucho más eficiente	10,00
Más eficiente	5,00

Igual	1,00
Menos eficiente	0,20
Mucho menos eficiente	0,10

Una vez definida la asignación del puntaje, se eligieron los parámetros para la selección del gestor de referencia como se muestra en la Tabla 2. Parámetros de comparación del gestor de referencia..

Tabla 2. Parámetros de comparación del gestor de referencia.

PARÁMETROS	DENOTACIÓN
Coste	A
Sistema operativo	B
Almacenamiento	C
Editor de formatos	D
Plugin para Word	E

Así mismo se realiza la comparación de los parámetros para asignarle un peso específico como se muestra en la Tabla 3. Peso de cada parámetro del gestor de referencia..

Tabla 3. Peso de cada parámetro del gestor de referencia.

PESO DE CADA PARÁMETRO							
	A	B	C	D	E	Suma	Factor
A	X	5,00	5,00	0,20	0,10	10,30	0,15
B	0,20	X	0,20	0,20	0,10	0,70	0,01
C	0,20	5,00	X	0,20	0,10	5,50	0,08
D	5,00	5,00	5,00	X	0,20	15,20	0,23
E	10,00	10,00	10,00	5,00	X	35,00	0,52
TOTAL						66,70	1,00

Como se muestra, el parámetro de mayor relevancia con un 52% fue plugin para Word, seguido con un 23% de editor de formato. Con valores más inferiores se encuentran coste, almacenamiento y sistema operativo con un 15%, 8% y 1% respectivamente.

Una vez asignados los pesos de cada parámetro, los gestores de referencia comparados fueron:

RefWorks, EndNote, Zotero, Mendeley y JabRef. A continuación, en la Tabla 4. Comparación del gestor de referencia respecto al coste (A). se observa comparación de los gestores de referencia respecto al parámetro coste, en la Tabla 5. Comparación del gestor de referencia respecto al sistema operativo (B). respecto al sistema operativo, en la Tabla 6. Comparación del gestor de referencia respecto al almacenamiento (C). respecto al almacenamiento, en la Tabla 7. Comparación del gestor de referencia respecto al editor de formatos (D). respecto editor de formatos y en la Tabla 8. Comparación del gestor de referencia respecto al plugin para Word (E). respecto al plugin para Word.

Tabla 4. Comparación del gestor de referencia respecto al coste (A).

COSTE (A)							
	RefWorks	EndNote	Zotero	Mendeley	JabRef	Suma	Factor
RefWorks	X	1.00	1.00	1.00	1.00	4.00	0.20
EndNote	1.00	X	1.00	1.00	1.00	4.00	0.20
Zotero	1.00	1.00	X	1.00	1.00	4.00	0.20
Mendeley	1.00	1.00	1.00	X	1.00	4.00	0.20
JabRef	1.00	1.00	1.00	1.00	X	4.00	0.20
TOTAL						20.00	1.00

Tabla 5. Comparación del gestor de referencia respecto al sistema operativo (B).

SISTEMA OPERATIVO (B)							
	RefWorks	EndNote	Zotero	Mendeley	JabRef	Suma	Factor
RefWorks	X	1.00	0.20	0.10	0.10	1.40	0.02147
EndNote	1.00	X	0.20	0.10	0.10	1.40	0.02147
Zotero	5.00	5.00	X	0.20	0.20	10.40	0.15951
Mendeley	10.00	10.00	5.00	X	1.00	26.00	0.39877
JabRef	10.00	10.00	5.00	1.00	X	26.00	0.39877
TOTAL						65.20	1.00

Tabla 6. Comparación del gestor de referencia respecto al almacenamiento (C).

ALMACENAMIENTO (C)							
	RefWorks	EndNote	Zotero	Mendeley	JabRef	Suma	Factor
RefWorks	X	5.00	10.00	1.00	1.00	17.00	0.29772

EndNote	0.20	X	5.00	0.20	0.20	5.60	0.09807
Zotero	0.10	0.20	X	0.10	0.10	0.50	0.00876
Mendeley	1.00	5.00	10.00	X	1.00	17.00	0.29772
JabRef	1.00	5.00	10.00	1.00	X	17.00	0.29772
TOTAL						57.10	1.00

Tabla 7. Comparación del gestor de referencia respecto al editor de formatos (D).

EDITOR DE FORMATOS (D)							
	RefWorks	EndNote	Zotero	Mendeley	JabRef	Suma	Factor
RefWorks	X	10.00	1.00	1.00	10.00	22.00	0.32070
EndNote	0.10	X	0.10	0.10	1.00	1.30	0.01895
Zotero	1.00	10.00	X	1.00	10.00	22.00	0.32070
Mendeley	1.00	10.00	1.00	X	10.00	22.00	0.32070
JabRef	0.10	1.00	0.10	0.10	X	1.30	0.01895
TOTAL						68.60	1.00

Tabla 8. Comparación del gestor de referencia respecto al plugin para Word (E).

PLUGIN PARA WORD (E)							
	RefWorks	EndNote	Zotero	Mendeley	JabRef	Suma	Factor
RefWorks	X	1.00	1.00	1.00	1.00	4.00	0.20
EndNote	1.00	X	1.00	1.00	1.00	4.00	0.20
Zotero	1.00	1.00	X	1.00	1.00	4.00	0.20
Mendeley	1.00	1.00	1.00	X	1.00	4.00	0.20
JabRef	1.00	1.00	1.00	1.00	X	4.00	0.20
TOTAL						20.00	1.00

Una vez realizada la comparación de los gestores de referencia respecto a cada parámetro seleccionado, se elaboró la matriz de decisión Tabla 9. Matriz de decisión de la selección del gestor de referencia., multiplicando el factor obtenido de cada comparación con el peso de cada parámetro de la Tabla 3. Peso de cada parámetro del gestor de referencia..

Tabla 9. Matriz de decisión de la selección del gestor de referencia.

DECISIÓN						
	A	B	C	D	E	Suma
RefWorks	0.0308846	0.0002253	0.0245499	0.073083	0.1049475	0.23

EndNote	0.0308846	0.0002253	0.008087	0.0043185	0.1049475	0.15
Zotero	0.0308846	0.001674	0.0007221	0.073083	0.1049475	0.21
Mendeley	0.0308846	0.004185	0.0245499	0.073083	0.1049475	0.24
JabRef	0.0308846	0.004185	0.0245499	0.0043185	0.1049475	0.17
TOTAL						1.00

Los resultados de la columna suma de la Tabla 9. Matriz de decisión de la selección del gestor de referencia. se representaron mediante un gráfico circular con el fin de apreciar mejor la diferencia porcentual entre los resultados individuales de cada gestor, ver Fig. 8. Representación gráfica de los resultados de la selección del gestor de referencia..

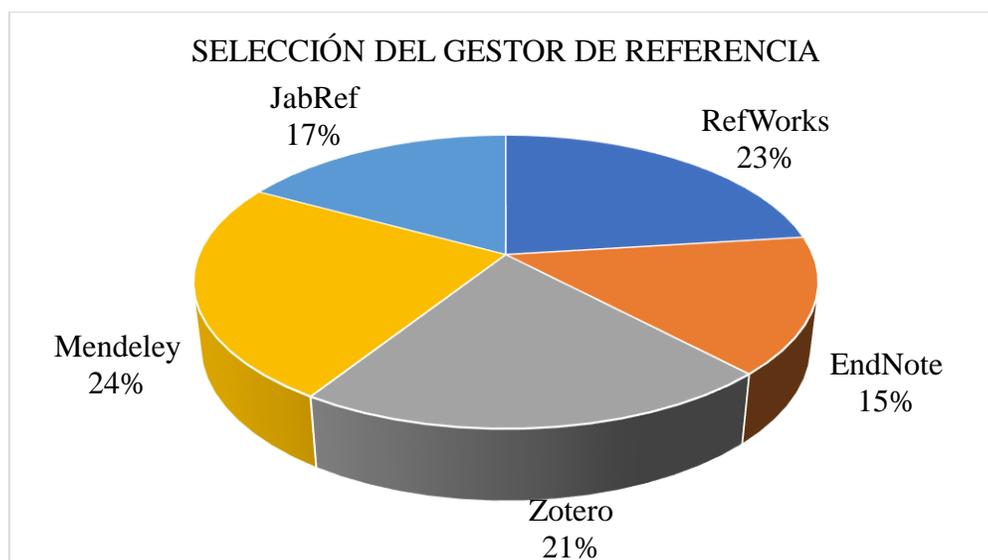


Fig. 8. Representación gráfica de los resultados de la selección del gestor de referencia.

Como se puede observar luego de utilizar la Matriz de Pugh el gestor de referencias que tuvo un mayor peso fue Mendeley con un 24%. Por lo tanto, fue seleccionado como el gestor de referencias a utilizar en el presente proyecto.

9.2. Selección de técnicas y herramientas de software

A continuación, se describirá la técnica de Inteligencia Artificial y herramienta de software seleccionado.

9.2.1. Técnicas de Inteligencia Artificial

Los modelos utilizados para la realización de predicciones se pueden resumir en dos grandes grupos: físicos y estocásticos. Los modelos físicos dependen de las características del terreno como rugosidad, altura, obstáculos y perfiles de viento. Ejemplos de modelos físicos son el análisis de imágenes en el cielo y la predicción numérica del clima. Por otro lado, los modelos estocásticos se basan en el análisis de un conjunto de datos y puede llevarse a cabo mediante varias técnicas como son:

- Series de Tiempo: Depende de los valores históricos de la misma variable que se desea predecir.
- Estadísticos: Trabajan con funciones estadísticas basadas en los datos históricos de la variable que se desea predecir y otras relacionadas.
- Inteligencia Artificial: Incluye la realización de modelos mediante algoritmos de Aprendizaje Automático (*Predicción de Potencia Eólica Utilizando Técnicas Modernas de Inteligencia Artificial*, n.d.).

Normalmente las Series de Tiempo y los métodos estadísticos no suelen ser tan precisos para realizar pronósticos a mediano y largo plazo por su dependencia del modelo matemático. Sin embargo, las técnicas más utilizadas de Inteligencia Artificial como las Redes Neuronales Artificiales, Redes Bayesianas y Máquinas de Soporte Vectorial se caracterizan por su capacidad de resolver problemas de gran dificultad y volumen. Con el fin de seleccionar objetivamente la técnica más apropiada para el presente proyecto se compararon estas últimas tres técnicas Red Neuronal Artificial, Red Bayesiana y Máquina de Soporte Vectorial mediante una matriz de Pugh. La asignación de puntaje se puede observar en la Tabla 1. Asignación de puntaje.. Los parámetros elegidos para la selección de la técnica de Inteligencia Artificial se observan en la Tabla 10.

Parámetros de comparación de la técnica de Inteligencia Artificial..

Tabla 10. Parámetros de comparación de la técnica de Inteligencia Artificial.

PARÁMETRO	DENOTACIÓN
Aplicaciones de predicción de tiempo	A
Habilidad con funciones no lineales	B
Cotas de error	C
Capacidad de entrenamiento	D

Así mismo se realiza la comparación de los parámetros para asignarle un peso específico como se muestra en la Tabla 11. Peso de cada parámetro de la técnica de Inteligencia Artificial..

Tabla 11. Peso de cada parámetro de la técnica de Inteligencia Artificial.

PESO DE CADA PARÁMETRO						
	A	B	C	D	Suma	Factor
A	X	1.00	5.00	10.00	16.00	0.42328
B	1.00	X	5.00	10.00	16.00	0.42328
C	0.20	0.20	X	5.00	5.40	0.14286
D	0.10	0.10	0.20	X	0.40	0.01058
TOTAL					37.80	1.00

A continuación, en la Tabla 12. Comparación de las técnicas de Inteligencia Artificial respecto a las aplicaciones de predicción de tiempo (A). se observa la comparación de las técnicas de Inteligencia Artificial respecto al parámetro aplicaciones de predicción de tiempo, en la Tabla 13. Comparación de las técnicas de Inteligencia Artificial respecto a la habilidad con funciones no lineales (B). respecto al parámetro habilidad con funciones no lineales, en la

Tabla 14. Comparación de las técnicas de Inteligencia Artificial respecto a cotas de error (C). respecto a cotas de error y en la Tabla 15. Comparación de las técnicas de Inteligencia Artificial respecto a la capacidad de entrenamiento (D). respecto a la capacidad de entrenamiento.

Tabla 12. Comparación de las técnicas de Inteligencia Artificial respecto a las aplicaciones de predicción de tiempo (A).

APLICACIONES DE PREDICCIÓN DE TIEMPO (A)

	Red Neuronal Artificial	Red Bayesiana	Máquina de Vector de Soporte	Suma	Factor
Red Neuronal Artificial	X	5.00	10.00	15.00	0.73171
Red Bayesiana	0.20	X	5.00	5.20	0.25366
Máquina de Vector de Soporte	0.10	0.20	X	0.30	0.01463
TOTAL				20.50	1.00

Tabla 13. Comparación de las técnicas de Inteligencia Artificial respecto a la habilidad con funciones no lineales (B).

HABILIDAD CON FUNCIONES NO LINEALES (B)					
	Red Neuronal Artificial	Red Bayesiana	Máquina de Vector de Soporte	Suma	Factor
Red Neuronal Artificial	X	5.00	5.00	10.00	0.80645
Red Bayesiana	0.20	X	1.00	1.20	0.09677
Máquina de Vector de Soporte	0.20	1.00	X	1.20	0.09677
TOTAL				12.40	1.00

Tabla 14. Comparación de las técnicas de Inteligencia Artificial respecto a cotas de error (C).

COTAS DE ERROR (C)					
	Red Neuronal Artificial	Red Bayesiana	Máquina de Vector de Soporte	Suma	Factor
Red Neuronal Artificial	X	5.00	5.00	10.00	0.80645
Red Bayesiana	0.20	X	1.00	1.20	0.09677
Máquina de Vector de Soporte	0.20	1.00	X	1.20	0.09677
TOTAL				12.40	1.00

Tabla 15. Comparación de las técnicas de Inteligencia Artificial respecto a la capacidad de entrenamiento (D).

CAPACIDAD DE ENTRENAMIENTO (D)					
	Red Neuronal Artificial	Red Bayesiana	Máquina de Vector de Soporte	Suma	Factor
Red Neuronal Artificial	X	0.10	0.10	0.20	0.00901
Red Bayesiana	10.00	X	1.00	11.00	0.49550
Máquina de Vector de Soporte	10.00	1.00	X	11.00	0.49550
TOTAL				22.20	1.00

Una vez realizada la comparación de las técnicas de Inteligencia Artificial respecto a cada parámetro seleccionado, se elaboró la matriz de decisión Tabla 16. Matriz de decisión de la selección de la técnica de Inteligencia Artificial. multiplicando el factor obtenido de cada comparación con el peso de cada parámetro de la Tabla 11. Peso de cada parámetro de la técnica de Inteligencia Artificial..

Tabla 16. Matriz de decisión de la selección de la técnica de Inteligencia Artificial.

DECISIÓN					
	A	B	C	D	Suma
Red Neuronal Artificial	0.30972	0.34136	0.11521	0.00010	0.76
Red Bayesiana	0.10737	0.04096	0.01382	0.00524	0.17
Máquina de Vector de Soporte	0.00619	0.04096	0.01382	0.00524	0.07
TOTAL					1.00

Por otra parte, los resultados de la columna suma de la Tabla 16. Matriz de decisión de la selección de la técnica de Inteligencia Artificial. se representaron mediante un gráfico circular con el fin de apreciar mejor la diferencia porcentual entre los resultados individuales de cada técnica. Ver Fig. 9. Representación gráfica de los resultados de la selección de la técnica de Inteligencia

Artificial.

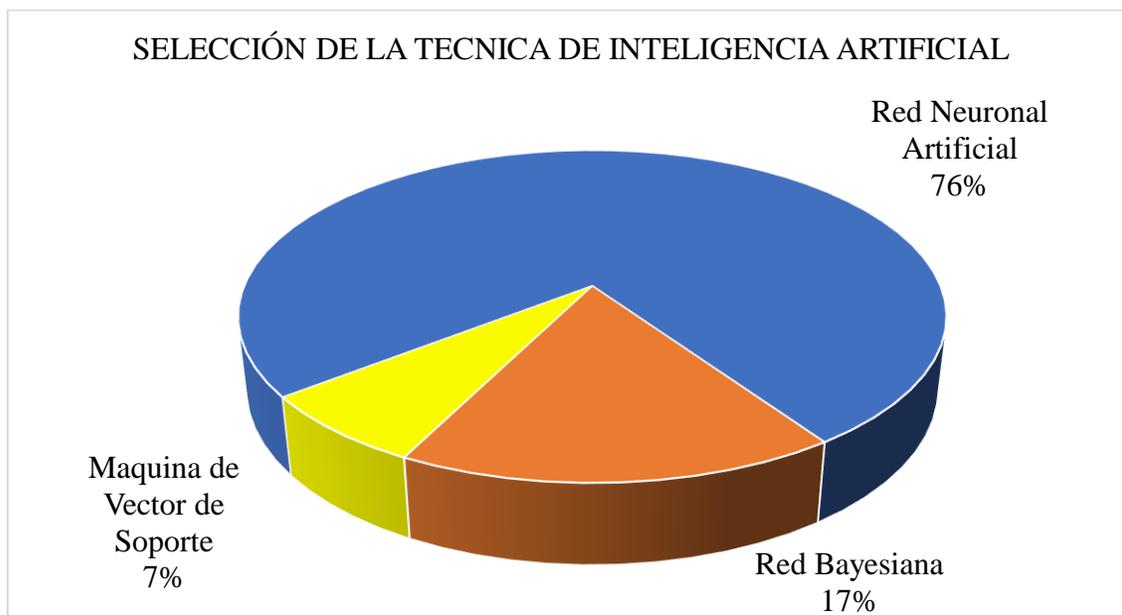


Fig. 9. Representación gráfica de los resultados de la selección de la técnica de Inteligencia Artificial.

A pesar de que pueda ser difícil comprender la naturaleza del funcionamiento interno de los modelos entrenados con Redes Neuronales Artificiales, ésta generalmente tiene mejores métricas de éxito comparado con otros métodos, lo que habla de su precisión tanto en el área de regresión como de clasificación. Otro indicativo que refuerza este argumento es el número de aplicaciones de tantos sectores de la industria en los que está ya inmerso el manejo de Redes Neuronales Artificiales, en contraste con campos más limitados de otros modelos. Su respuesta precisa ante funciones de salida no lineales lo hace ideal realizar predicciones del Recurso Eólico disponible debido a la gran variabilidad que presentan los datos meteorológicos como la velocidad del viento y temperatura. Todas las ventajas explicadas anteriormente se ven reflejadas en la Fig. 9. Representación gráfica de los resultados de la selección de la técnica de Inteligencia Artificial. luego de realizar la matriz de Pugh, donde la técnica Red Neuronal Artificial logra un 76%, es por ello que para el presente proyecto se seleccionó las Redes Neuronales Artificiales como técnica para el uso de la tecnología de Inteligencia Artificial.

9.2.2. Selección del lenguaje de programación

En vista de que hoy en día son muchos los lenguajes de programación empleados en el uso de la Ciencia de los Datos y el Aprendizaje Automático, se reunió la información más relevante de cada uno de ellos (Rojas, n.d.) y se seleccionó mediante una matriz de Pugh el que más se adaptaba a las necesidades del proyecto. En la Tabla 1. Asignación de puntaje. se observa la asignación de puntaje, en la Tabla 17. Parámetros de comparación del lenguaje de programación. los parámetros elegidos para la comparación con los lenguajes de programación: Python, R, Matlab y Julia.

Tabla 17. Parámetros de comparación del lenguaje de programación.

PARÁMETROS	DENOTACIÓN
Sistema operativo	A
Tipo de licencia	B
Disponibilidad de librerías	C
Especialización del software	D
Interpretación en entornos de desarrollo	E

A continuación, se realiza la comparación de los parámetros para asignarle un peso específico como se muestra en la Tabla 18. Peso de cada parámetro del lenguaje de programación..

Tabla 18. Peso de cada parámetro del lenguaje de programación.

PESO DE CADA PARÁMETRO							
	A	B	C	D	E	Suma	Factor
A	X	1.00	0.10	0.10	0.10	1.30	0.01662
B	1.00	X	0.10	0.10	0.10	1.30	0.01662
C	10.00	10.00	X	5.00	5.00	30.00	0.38363
D	10.00	10.00	0.20	X	0.20	20.40	0.26087
E	10.00	10.00	0.20	5.00	X	25.20	0.32225
TOTAL						78.20	1.00

Una vez asignados los pesos de cada parámetro, se prosiguió a comparar cada lenguaje respecto a cada uno de los parámetros, en la Tabla 19. Comparación del lenguaje de programación respecto al sistema operativo (A). se observa la comparación respecto al parámetro sistema operativo, en la

Tabla 20. Comparación del lenguaje de programación respecto al tipo de licencia (B). respecto al tipo de licencia, en la Tabla 21. Comparación del lenguaje de programación respecto a la disponibilidad de librerías (C). respecto a la disponibilidad de librerías, en la Tabla 22. Comparación del lenguaje de programación respecto a la especialización del software (D). respecto a la especialización del software y en la Tabla 23. Comparación del lenguaje de programación respecto a la interpretación en entornos de desarrollo (E). respecto a la interpretación de entornos de desarrollo.

Tabla 19. Comparación del lenguaje de programación respecto al sistema operativo (A).

SISTEMA OPERATIVO (A)						
	Python	R	Matlab	Julia	Suma	Factor
Python	X	1.00	1.00	1.00	3.00	0.25
R	1.00	X	1.00	1.00	3.00	0.25
Matlab	1.00	1.00	X	1.00	3.00	0.25
Julia	1.00	1.00	1.00	X	3.00	0.25
TOTAL					12.00	1.00

Tabla 20. Comparación del lenguaje de programación respecto al tipo de licencia (B).

TIPO DE LICENCIA (B)						
	Python	R	Matlab	Julia	Suma	Factor
Python	X	5.00	10.00	1.00	16.00	0.37471
R	0.20	X	10.00	0.20	10.40	0.24356
Matlab	0.10	0.10	X	0.10	0.30	0.00703
Julia	1.00	5.00	10.00	X	16.00	0.37471
TOTAL					42.70	1.00

Tabla 21. Comparación del lenguaje de programación respecto a la disponibilidad de librerías (C).

DISPONIBILIDAD DE LIBRERIAS (C)						
	Python	R	Matlab	Julia	Suma	Factor
Python	X	5.00	10.00	5.00	20.00	0.60790
R	0.20	X	5.00	1.00	6.20	0.18845
Matlab	0.10	0.20	X	0.20	0.50	0.01520
Julia	0.20	1.00	5.00	X	6.20	0.18845

TOTAL	32.90	1.00
-------	-------	------

Tabla 22. Comparación del lenguaje de programación respecto a la especialización del software (D).

ESPECIALIZACIÓN DEL SOFTWARE (D)						
	Python	R	Matlab	Julia	Suma	Factor
Python	X	5.00	5.00	5.00	15.00	0.53571
R	0.20	X	5.00	1.00	6.20	0.22143
Matlab	0.20	0.20	X	0.20	0.60	0.02143
Julia	0.20	1.00	5.00	X	6.20	0.22143
TOTAL					28.00	1.00

Tabla 23. Comparación del lenguaje de programación respecto a la interpretación en entornos de desarrollo (E).

INTERPRETACIÓN EN ENTORNOS DE DESARROLLO (E)						
	Python	R	Matlab	Julia	Suma	Factor
Python	X	5.00	0.20	5.00	10.20	0.26984
R	0.20	X	0.10	1.00	1.30	0.03439
Matlab	5.00	10.00	X	10.00	25.00	0.66138
Julia	0.20	1.00	0.10	X	1.30	0.03439
TOTAL					37.80	1.00

Una vez realizada la comparación de los lenguajes de programación respecto a cada parámetro seleccionado, se elaboró la matriz de decisión Tabla 24. Matriz de decisión de la selección del lenguaje de programación. multiplicando el factor obtenido de cada comparación con el peso de cada parámetro de la Tabla 18. Peso de cada parámetro del lenguaje de programación..

Tabla 24. Matriz de decisión de la selección del lenguaje de programación.

DECISIÓN						
	A	B	C	D	E	Suma
Python	0.00416	0.00623	0.23321	0.13975	0.08696	0.47
R	0.00416	0.00405	0.07230	0.05776	0.01108	0.15
Matlab	0.00416	0.00012	0.00583	0.00559	0.21313	0.23
Julia	0.00416	0.00623	0.07230	0.05776	0.01108	0.15
TOTAL						1.00

Por otra parte, los resultados de la columna suma de la Tabla 24. Matriz de decisión de la selección del lenguaje de programación. se representaron mediante un gráfico circular con el fin de apreciar mejor la diferencia porcentual entre los resultados individuales de cada lenguaje de programación. Ver Fig. 10. Representación gráfica de los resultados de la selección del lenguaje de programación..

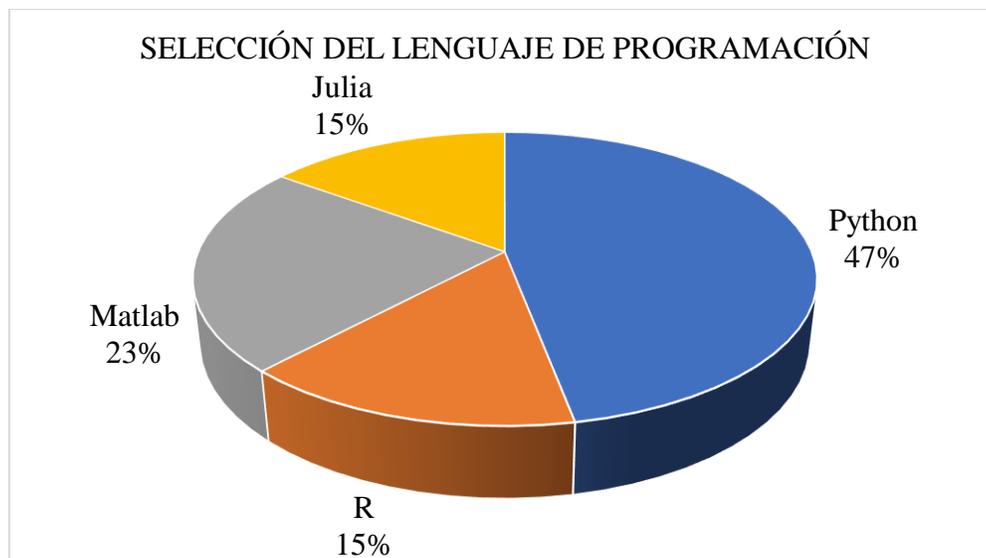


Fig. 10. Representación gráfica de los resultados de la selección del lenguaje de programación.

Como se puede observar luego de utilizar la matriz de Pugh el lenguaje de programación que tuvo un mayor peso en la comparación con los parámetros seleccionados fue Python con un 47%. Por lo tanto, se ha seleccionado como el lenguaje de programación a utilizar en el presente proyecto.

9.2.3. Software y entornos virtuales

A pesar de que es muy común seleccionar el software o el entorno que más se adapte a las necesidades de un proyecto, en esta ocasión se creyó conveniente utilizar distintos entornos según las necesidades de cada etapa independiente del proyecto, beneficiándose de esta forma de las ventajas de cada uno de ellos. Los entornos de programación utilizados fueron Google Colaboratory y Visual Studio Code. A continuación, se muestra una breve descripción de cada uno

de ellos, las ventajas de su uso, así como la etapa en que fueron utilizados.

A. Google Colaboratory

Es un producto de Google Research y permite desarrollar código en Python en el navegador. Es útil para desarrollar aplicaciones de aprendizaje automático, análisis de datos y educación, y es que Google Colaboratory permite la entrada de texto lo que hace más dinámico su uso. Este entorno no requiere instalaciones de paquetes ni configuraciones. Además, cuenta con una GPU y otros recursos computacionales gratuitos (*Google Research*, n.d.). Este entorno fue utilizado para la etapa de preprocesamiento y exploración de los datos, así como el entrenamiento y la evaluación del modelo. La ventana de contar con una GPU permitió acelerar significativamente los procesos de entrenamiento del algoritmo de la Red Neuronal Artificial, lo cual fue de suma importancia teniendo en cuenta que se utilizaron cerca de 900.000 datos.

B. Visual Studio Code

Es un entorno de código abierto y multiplataforma desarrollado por Microsoft, el cual incluye soporte para JavaScript, HTML, y CSS. Además, soporta lenguajes adicionales, como Python, por medio de extensiones (April Speight, 2021). Dentro de los recursos de Python disponibles en Visual Studio Code se encuentra el paquete Tkinter, el cual es el más utilizado para el desarrollo de interfaces caracterizado por su fácil uso y manejo. Teniendo en cuenta las características mencionadas anteriormente se utilizó este entorno para la realización de la interfaz gráfica en la cual se presentarán los resultados de las predicciones obtenidas en el entorno de Google Colaboratory.

9.3. Creación del modelo de Ciencia de los Datos

Para la construcción de un modelo de Ciencia de los Datos, específicamente de Aprendizaje Automático es necesario utilizar una serie de pasos que permitan garantizar el funcionamiento

óptimo de las predicciones que en este caso se desean realizar. En la Fig. 11. Proceso de creación de un modelo de Aprendizaje Automático. se puede observar el paso a paso de la construcción de un modelo de este tipo, y a continuación el desarrollo de cada uno de ellos.

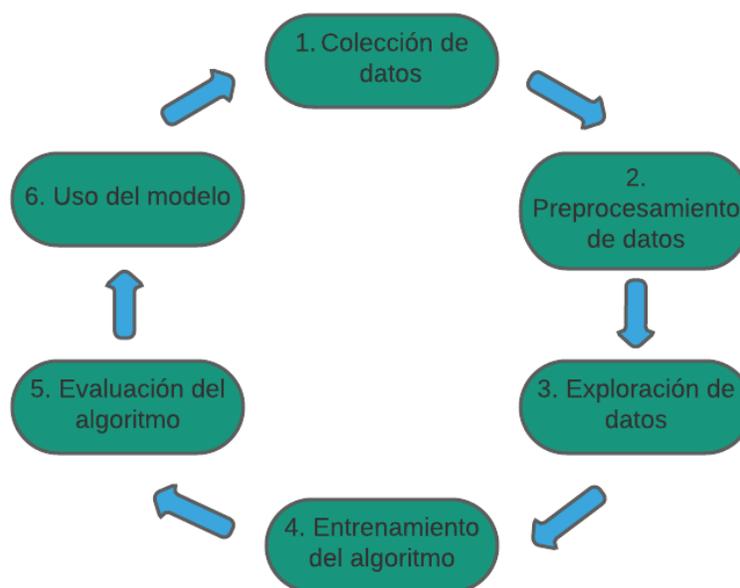


Fig. 11. Proceso de creación de un modelo de Aprendizaje Automático. (Rojas, n.d.).

9.3.1. Colección de datos

El principal paso para realizar un modelo predictivo es la recolección de los datos que permitirán entrenar el modelo. Los datos usualmente suelen obtenerse de una página web, una API o una base de datos (Rojas, n.d.). En este caso se utilizó una base de datos de velocidad de viento obtenidas del IDEAM en la estación meteorológica del Aeropuerto Camilo Daza, en un rango de 4 años y una frecuencia de 2 minutos, para un total de 890145 datos. También se obtuvieron los valores de temperatura promedio anual para los 4 años. Todos los datos correspondientes a la estación meteorológica corresponden a una altura de 10 m. En la Fig. 12. Bases de datos cargadas y definidas mediante un marco de datos en Python. se encuentra el fragmento de código en Python con el que se cargaron las bases de datos de formato csv, y la definición del marco de datos para

cada una, con el fin de manipular y usar la información cargada al entorno de Google Colaboratoy. En la Fig. 13. Marco de datos correspondiente a los valores de velocidad de viento para un rango de 4 años y una frecuencia de 2 minutos. se puede observar el marco de datos correspondientes a los valores de velocidad de viento y en la Fig. 14. Marco de datos correspondiente a los valores de velocidad de viento para un rango de 4 años y una frecuencia de 2 minutos. el marco de datos de la temperatura, ambos definidos y cargados en Python.

```
#Rutas de bases de datos en Google Drive
path1='/content/drive/MyDrive/PROYECTO DE GRADO/BASE DE DATOS/DATA.csv'
path2='/content/drive/MyDrive/PROYECTO DE GRADO/BASE DE DATOS/Temp.csv'

#Definir dataframes con pandas

df1=pd.read_csv(path1)
df2=pd.read_csv(path2)
```

Fig. 12. Bases de datos cargadas y definidas mediante un marco de datos en Python.

	Fecha	VelocidadDeViento
0	01/01/2017 00:00	2.2
1	01/01/2017 00:02	2.1
2	01/01/2017 00:04	2.3
3	01/01/2017 00:06	2.1
4	01/01/2017 00:08	2
...
891473	31/12/2020 22:54	1.4
891474	31/12/2020 22:56	1.2
891475	31/12/2020 22:58	1.6
891476	31/12/2020 23:00	1.1
891477	NaN	NaN
891478 rows x 2 columns		

Fig. 13. Marco de datos correspondiente a los valores de velocidad de viento para un rango de 4 años y una frecuencia de 2 minutos.

	Fecha	Valor
0	01/01/2017 00:00	27.354348
1	01/01/2018 00:00	26.831481
2	01/01/2019 00:00	27.679116
3	01/01/2020 00:00	27.290892

Fig. 14. Marco de datos correspondiente a los valores de velocidad de viento para un rango de 4 años y una frecuencia de 2 minutos.

9.3.2. Preprocesamiento de los datos

Una vez se tengan los datos es necesario realizar las tareas necesarias para asegurarse de que todos los datos se encuentren en el mismo formato y el algoritmo se entrene correctamente. Diferentes condiciones y sentencias fueron ejecutadas en Python para asegurarse de que el formato fuera igual para todas las variables. Para este caso se tienen dos marcos de datos, uno de velocidad de viento y uno de temperatura, ambos con la variable fecha. Para establecer condiciones con respecto a las variables es necesario que todas se encuentren en el formato de tipo numérico. Para ello de la variable “Fecha” que es de tipo “datetime64” en el marco de datos de velocidad de viento, se ha extraído las variables numéricas Año, Mes, Día, Hora, Minuto, ver Fig. 15. Conversión del tipo de dato datetime a numérico para la variable Fecha en el marco de datos de velocidad.. Así como la conversión del tipo de dato de la variable velocidad de viento a numérico, el cual era inicialmente de tipo objeto, ver Fig. 16. Conversión del tipo de dato objeto a numérico para velocidad de viento en el marco de datos de velocidad.. Para el marco de datos de temperatura se realizó el mismo procedimiento para la variable “Fecha”, ver Fig. 17. Conversión del tipo de dato

datetime a numérico para la variable Fecha en el marco de datos de temperatura.. Mientras que para la variable temperatura no fue necesario cambiar el tipo de dato ya que estaba establecida como tipo numérica, sin embargo, se pasó de grados Celsius a Kelvin para facilitar cálculos posteriores. Finalmente, en la Fig. 18. Marco de datos de tipo numérico con las entradas del modelo. se puede apreciar como ha quedado el marco de datos con las variables de entrada del modelo una vez convertido todas las variables a tipo numérico y unidos en un mismo marco de datos, donde la temperatura está expresada en Kelvin y la velocidad de viento en (m/s).

```
df_WV['Fecha'] = pd.to_datetime(df_WV['Fecha'], errors='coerce', format='%d/%m/%Y %H:%M')
df_WV['Año'] = df_WV['Fecha'].dt.year
df_WV['Mes'] = df_WV['Fecha'].dt.month
df_WV['Dia'] = df_WV['Fecha'].dt.day
df_WV['Hora'] = df_WV['Fecha'].dt.hour
df_WV['Minuto'] = df_WV['Fecha'].dt.minute
```

Fig. 15. Conversión del tipo de dato datetime a numérico para la variable Fecha en el marco de datos de velocidad.

```
[ ] df_WV.VelocidadDeViento = pd.to_numeric(df_WV.VelocidadDeViento, errors='coerce')
```

Fig. 16. Conversión del tipo de dato objeto a numérico para velocidad de viento en el marco de datos de velocidad.

```
df_T['Fecha'] = df_T['Fecha'].astype('datetime64[ns]')
df_T['Año'] = df_T['Fecha'].dt.year
df_T['Mes'] = df_T['Fecha'].dt.month
df_T['Dia'] = df_T['Fecha'].dt.day
df_T['Hora'] = df_T['Fecha'].dt.hour
df_T['Minuto'] = df_T['Fecha'].dt.minute
```

Fig. 17. Conversión del tipo de dato datetime a numérico para la variable Fecha en el marco de datos de temperatura.

	Año	Mes	Día	Hora	Minuto	TemperaturaK	VelocidadDeViento
0	2017.0	1.0	1.0	0.0	0.0	300.504	2.2
1	2017.0	1.0	1.0	0.0	2.0	300.504	2.1
2	2017.0	1.0	1.0	0.0	4.0	300.504	2.3
3	2017.0	1.0	1.0	0.0	6.0	300.504	2.1
4	2017.0	1.0	1.0	0.0	8.0	300.504	2.0
...
891472	2020.0	12.0	31.0	22.0	52.0	300.441	2.0
891473	2020.0	12.0	31.0	22.0	54.0	300.441	1.4
891474	2020.0	12.0	31.0	22.0	56.0	300.441	1.2
891475	2020.0	12.0	31.0	22.0	58.0	300.441	1.6
891476	2020.0	12.0	31.0	23.0	0.0	300.441	1.1

890145 rows × 7 columns

Fig. 18. Marco de datos de tipo numérico con las entradas del modelo.

9.3.3. Exploración de los datos

En la exploración de los datos se analizan si existen valores faltantes o atípicos que pudieran afectar al modelo. Además, también se busca si existe algún tipo de patrón que sirva de ayuda a la construcción del mismo. En la Fig. 19. Resumen estadístico del marco de datos. se puede observar un resumen estadístico de los datos, como valores promedio, mínimos, máximos, así como los percentiles 25,50 y 75. Esto fue de mucha utilidad para observar que a pesar de que el 75% de los datos de velocidad de viento corresponden a un valor de 5.7 (m/s) o menor, y el valor máximo es 999.9 (m/s), lo cual es un valor atípico al ser excesivamente alto. Debido a que los valores que superaban los 15 (m/s) eran escasos y muy distantes entre ellos se acotó el rango de datos de la variable velocidad de viento a un máximo de 15 (m/s) para evitar malos pronósticos al usar el modelo. La instrucción utilizada para acotar el rango, así como el resumen estadístico del nuevo marco de datos se presentan en la Fig. 20. Acote del rango y resumen estadístico del nuevo marco de datos..

	Año	Mes	Día	Hora	Minuto	VelocidadDeViento
count	891430.000000	891430.000000	891430.000000	891430.000000	891430.000000	891430.000000
mean	2018.435064	6.467597	15.847223	11.259040	29.038215	4.415768
std	1.096977	3.468874	8.764593	6.813686	17.327244	23.746525
min	2017.000000	1.000000	1.000000	0.000000	0.000000	0.000000
25%	2017.000000	3.000000	8.000000	5.000000	14.000000	1.400000
50%	2018.000000	7.000000	16.000000	11.000000	30.000000	2.900000
75%	2019.000000	9.000000	23.000000	17.000000	44.000000	5.700000
max	2020.000000	12.000000	31.000000	23.000000	58.000000	999.900000

Fig. 19. Resumen estadístico del marco de datos.

```
df_vw=df_vw[df_vw['VelocidadDeViento']<=15]
```

```
df_vw.describe()
```

	Año	Mes	Día	Hora	Minuto	VelocidadDeViento
count	890145.000000	890145.000000	890145.000000	890145.000000	890145.000000	890145.000000
mean	2018.434592	6.467218	15.848755	11.258759	29.038408	3.847193
std	1.097080	3.469636	8.764616	6.817204	17.327013	2.980743
min	2017.000000	1.000000	1.000000	0.000000	0.000000	0.000000
25%	2017.000000	3.000000	8.000000	5.000000	14.000000	1.400000
50%	2018.000000	7.000000	16.000000	11.000000	30.000000	2.900000
75%	2019.000000	9.000000	23.000000	17.000000	44.000000	5.700000
max	2020.000000	12.000000	31.000000	23.000000	58.000000	15.000000

Fig. 20. Acote del rango y resumen estadístico del nuevo marco de datos.

Por otra parte, para tener un conjunto de datos limpios también es necesario verificar si existen datos faltantes o nulos, es por esta razón que se utiliza la instrucción que se muestra en la Fig. 21. Conteo de valores nulos por columnas.. La cual realiza el conteo de valores nulos por columnas, en esta ocasión se verificó que no existía ningún valor nulo.

```
#Conteo de valores nulos x columnas
df_vw.isnull().sum().sum()
```

0

Fig. 21. Conteo de valores nulos por columnas.

Otro análisis que fue necesario realizar con los datos limpios antes del entrenamiento fue el de la variación de la velocidad media a largo plazo con la altura, lo que se conoce como perfil vertical

de viento. En este caso se estima esta variación a partir de los valores promedio mensuales de velocidad de viento a una altura específica mediante la ley de Hellman de la ecuación (7). En la Fig. 22. Perfil de viento para la zona del Aeropuerto. se puede observar la gráfica del perfil de viento para la zona del Aeropuerto.

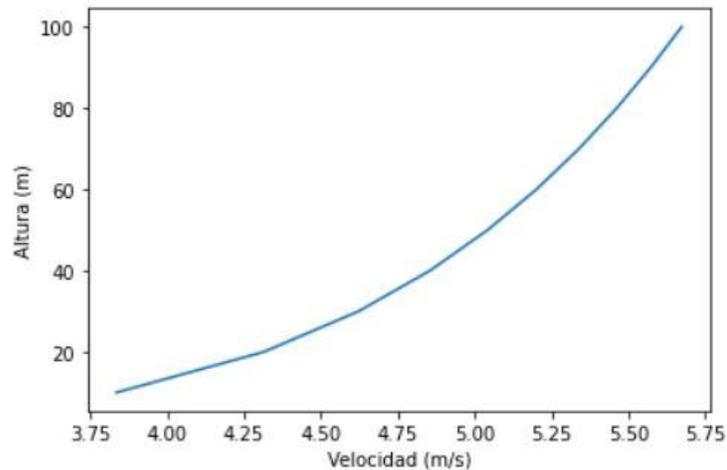


Fig. 22. Perfil de viento para la zona del Aeropuerto.

Además, se calculó la potencia disponible por área mediante la ecuación (6). Con los parámetros constantes de presión a nivel del mar P_o , gravedad g , altura sobre el nivel del mar Z , la constante universal sobre masa molecular del gas para el aire R_p siguientes:

$$P_o = 101250 [Pa]$$

$$Z = 313 [m]$$

$$R_p = 287 \left[\frac{J}{kg \cdot K} \right]$$

$$g = 9.8 \left[\frac{m}{s} \right]$$

Los valores de T_k y v , corresponden a los valores de temperatura y velocidad de viento del marco de datos preprocesado del apartado anterior. Los valores calculados de densidad de potencia fueron agregados al marco de datos y puede apreciarse en la Fig. 23. Marco de datos con las entradas y salida del modelo. donde la densidad de potencia tiene unidades de W/m^2 . A su vez,

en la Fig. 24. Resumen estadístico del marco de datos. puede observarse el resumen estadístico de todas las variables del marco de datos.

	Año	Mes	Día	Hora	Minuto	TemperaturaK	VelocidadDeViento	DensidadDePotencia
0	2017.0	1.0	1.0	0.0	0.0	300.504348	2.2	6.027105
1	2017.0	1.0	1.0	0.0	2.0	300.504348	2.1	5.242019
2	2017.0	1.0	1.0	0.0	4.0	300.504348	2.3	6.886907
3	2017.0	1.0	1.0	0.0	6.0	300.504348	2.1	5.242019
4	2017.0	1.0	1.0	0.0	8.0	300.504348	2.0	4.528253
...
891472	2020.0	12.0	31.0	22.0	52.0	300.440892	2.0	4.529175
891473	2020.0	12.0	31.0	22.0	54.0	300.440892	1.4	1.553507
891474	2020.0	12.0	31.0	22.0	56.0	300.440892	1.2	0.978302
891475	2020.0	12.0	31.0	22.0	58.0	300.440892	1.6	2.318938
891476	2020.0	12.0	31.0	23.0	0.0	300.440892	1.1	0.753541

890145 rows x 8 columns

Fig. 23. Marco de datos con las entradas y salida del modelo.

	Año	Mes	Día	Hora	Minuto	TemperaturaK	VelocidadDeViento	DensidadDePotencia
count	890145.000000	890145.000000	890145.000000	890145.000000	890145.000000	890145.000000	890145.000000	890145.000000
mean	2018.434592	6.467218	15.848755	11.258759	29.038408	300.442674	3.847193	104.989665
std	1.097080	3.469636	8.764616	6.817204	17.327013	0.308301	2.980743	208.213788
min	2017.000000	1.000000	1.000000	0.000000	0.000000	299.981481	0.000000	0.000000
25%	2017.000000	3.000000	8.000000	5.000000	14.000000	299.981481	1.400000	1.557040
50%	2018.000000	7.000000	16.000000	11.000000	30.000000	300.504348	2.900000	13.839154
75%	2019.000000	9.000000	23.000000	17.000000	44.000000	300.829116	5.700000	104.908513
max	2020.000000	12.000000	31.000000	23.000000	58.000000	300.829116	15.000000	1915.090642

Fig. 24. Resumen estadístico del marco de datos.

9.3.4. Entrenamiento del modelo

Una vez se ha limpiado el marco de datos en las etapas de preprocesamiento y exploración se puede proseguir con el entrenamiento del modelo, para el cual se especifican a continuación, los parámetros como el número de entradas, salidas, partición de los datos, algoritmo de entrenamiento, cantidad de épocas y métrica de éxito del modelo utilizado.

9.3.5. Parámetros

En este caso las entradas del modelo fueron las variables de Año, Mes, Día, Hora, Minuto, Temperatura y Velocidad de viento, para un total de 7 entradas. Mientras que la salida corresponde

a la variable Densidad de potencia, es decir, una salida. Con respecto a la partición de los datos se tomaron un 78,4% para entrenamiento y un 21,6% para evaluación ya que fue la partición que generaba un año de pronóstico. A pesar de que en teoría 1 año sería el 25% de los 4 años totales, también es cierto que no se tiene la misma cantidad de datos por año. Con respecto al algoritmo de entrenamiento se utilizó el método de retropropagación y la técnica de validación cruzada para evaluar los resultados del modelo, garantizando la independencia entre datos de entrenamiento y validación. Para el entrenamiento existen otros parámetros como el número de épocas, capas y neuronas. Estos últimos se han variado de tal forma de seleccionar el que resultara en un error de predicción más pequeño, el cual se alcanzó para 300 épocas, 1 capa oculta, 64 neuronas para la capa oculta y la capa de entrada, y una neurona para la capa de salida. Por último, la métrica de éxito seleccionada para evaluar el modelo fue el Error Absoluto Promedio EAE o MAE del inglés Mean Absolute Error ya que es la métrica más utilizada y recomendada para los modelos de regresión. En la Tabla 25. Resumen de los parámetros utilizados para el entrenamiento del modelo. se puede observar el resumen de los parámetros del modelo utilizado.

Tabla 25. Resumen de los parámetros utilizados para el entrenamiento del modelo.

Número de entradas	Número de salidas	Partición de los datos	Algoritmo de entrenamiento	Número de capas ocultas	Neuronas en la capa oculta	Número de épocas	Métrica de éxito
7	1	78.4% entrenamiento 21.6% validación	Retropropagación	1	64	300	EAM

9.3.6. Evaluación del modelo

Para la evaluación del modelo se tuvo en cuenta la variación del número de épocas con el fin de aumentar la precisión, es decir que el MAE disminuyera lo más posible. Se pudo observar que a medida que se aumentaba el número de épocas, el error disminuía, sin embargo, para un número de épocas mayor se requiere un tiempo de simulación mayor, lo que puede generar un gasto considerablemente mayor de tiempo. Por esta razón se fijó el modelo para 300 épocas. En la Tabla 26. Relación entre el número de épocas y el MAE. se puede apreciar los resultados del error MAE conforme aumentaban las épocas. Por otra parte, en la Fig. 25. Relación entre el número de épocas y el MAE para la zona de Aeropuerto. se puede observar la relación entre el número de épocas y el MAE. Además, una comparación entre los valores reales y los valores predichos promedios mensuales para 1 año se puede apreciar en la Fig. 26. Valores reales vs valores predichos mensuales para el año 2020 zona Aeropuerto..

Tabla 26. Relación entre el número de épocas y el MAE.

Número de épocas	MAE
50	21.2163
100	4.0124
150	2.9001
200	2.3458
250	2.2038
300	1.6449

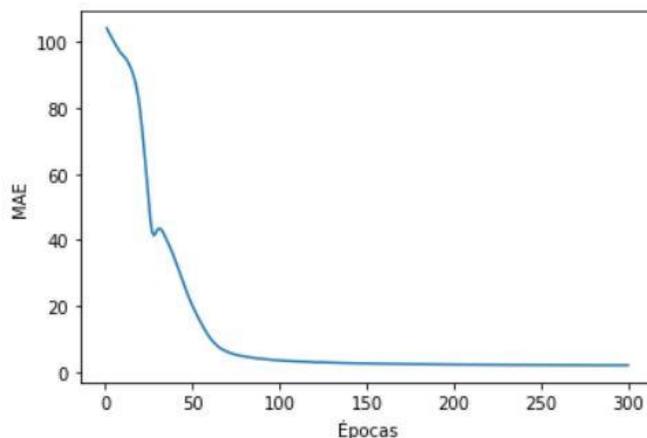


Fig. 25. Relación entre el número de épocas y el MAE para la zona de Aeropuerto.

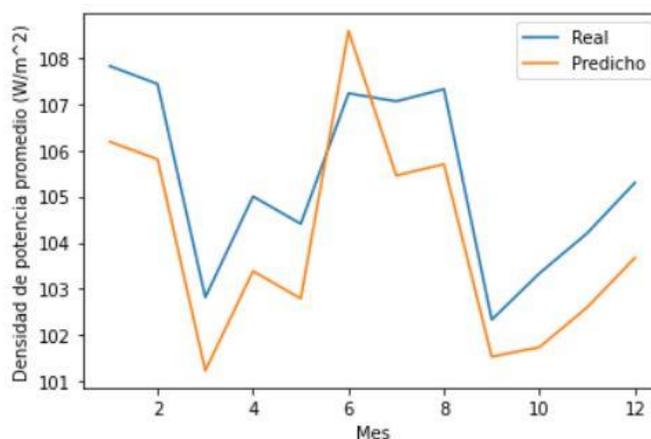


Fig. 26. Valores reales vs valores predichos mensuales para el año 2020 zona Aeropuerto.

9.3.7. Uso del modelo

El uso del modelo una vez se comprobó su buena precisión se aplicó para realizar predicciones en 3 zonas distintas de la ciudad de Cúcuta: Ceiba, Torcoroma y Belén. Los análisis correspondientes a estas zonas se encuentran en el apartado siguiente de pruebas de funcionamiento sobre datos experimentales.

9.3.8. Pruebas de funcionamiento sobre datos experimentales

Una vez evaluado el modelo se procedió a tomar un conjunto de datos experimentales en 3 zonas de la ciudad: Ceiba, Torcoroma y Belén. Con el fin de aplicar el algoritmo desarrollado y realizar una comparativa de las predicciones del Recurso Eólico en las zonas mencionadas. Para

ello, se llevó a cabo la misma metodología realizada en el apartado anterior. A continuación, se muestran los detalles del paso a paso.

A. Colección de datos

Para la colección de datos experimentales se utilizó un anemómetro de álabe con el fin de capturar los datos de velocidad de viento y temperatura en las zonas mencionadas previamente. En la Fig. 27. Anemómetro fijado en la zona de Torcoroma. se puede observar el anemómetro utilizado.

Para la zona Ceiba se lograron capturar un total de 10171 datos en un tiempo de 1 hora y 25 minutos, para la zona de Torcoroma se obtuvieron 10006 mediciones en un tiempo de 1 hora y 5 minutos y en la zona de Belén se tomaron 7770 datos en un tiempo también de 1 hora y 5 minutos. Todos los datos fueron tomados a una altura de 3 metros. En las Fig. 28. Marco de datos en Python para la zona de Ceiba., Fig. 29. Marco de datos en Python para la zona de Torcoroma. y Fig. 30. Marco de datos en Python para la zona de Belén. se muestran los marcos de datos para Ceiba, Torcoroma y Belén respectivamente.



Fig. 27. Anemómetro fijado en la zona de Torcoroma.

	Time	TemperaturaK	VelocidadDeViento
12 10 2021	16:41:24	305.45	0.0
12 10 2021	16:41:24	305.45	0.0
12 10 2021	16:41:25	305.45	0.0
12 10 2021	16:41:25	305.45	0.0
12 10 2021	16:41:26	305.45	0.0
...
12 10 2021	18:06:11	302.55	1.0
12 10 2021	18:06:12	302.55	1.0
12 10 2021	18:06:12	302.55	1.0
12 10 2021	18:06:13	302.55	1.0
12 10 2021	18:06:13	302.55	1.0

10171 rows x 3 columns

Fig. 28. Marco de datos en Python para la zona de Ceiba.

	Time	TemperaturaK	VelocidadDeViento
07 10 2021	15:23:32	311.35	0.0
07 10 2021	15:23:32	311.35	0.0
07 10 2021	15:23:33	311.35	0.0
07 10 2021	15:23:33	311.35	0.0
07 10 2021	15:23:34	311.35	0.0
...
07 10 2021	16:46:59	305.05	1.7
07 10 2021	16:47:00	305.05	1.9
07 10 2021	16:47:00	305.05	2.2
07 10 2021	16:47:01	305.05	2.2
07 10 2021	16:47:01	305.05	2.2

10006 rows x 3 columns

Fig. 29. Marco de datos en Python para la zona de Torcoroma.

	Time	T	V
15 10 2021	21:19:47	25.7	1.8
15 10 2021	21:19:48	25.7	1.8
15 10 2021	21:19:48	25.7	1.6
15 10 2021	21:19:49	25.7	1.6
15 10 2021	21:19:49	25.7	1.6
...
15 10 2021	22:24:34	24.6	0.0
15 10 2021	22:24:35	24.6	0.0
15 10 2021	22:24:35	24.6	0.0
15 10 2021	22:24:36	24.6	0.0
15 10 2021	22:24:36	24.6	0.0

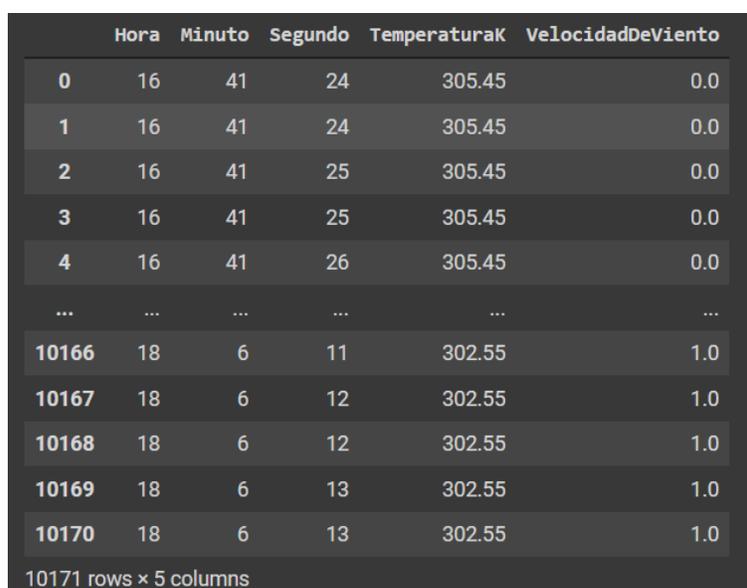
7770 rows × 3 columns

Fig. 30. Marco de datos en Python para la zona de Belén.

B. Preprocesamiento

La diferencia de altura entre los datos tomados de la estación meteorológica y los obtenidos experimentalmente permite que éstos últimos tiendan a tener menos valores atípicos muy altos ya que se encuentran a una altura mucho menor. En este sentido no fue necesario acotar el rango de valores obtenidos. Tampoco se detectaron valores nulos o faltantes, lo que permitió agilizar gran parte del preprocesamiento. El formato con que se exportaron los datos fue compatible con los utilizados por Python. Por lo tanto, sólo fue necesario llevar todas las variables capturadas a datos

tipo numérico, con el fin de poder prepararlos para el entrenamiento del modelo. Las instrucciones que se utilizaron para realizar la conversión del tipo de datos son las mismas que para la zona del aeropuerto. En las Fig. 31. Marco de datos de tipo numérico con las entradas del modelo para Ceiba., Fig. 32. Marco de datos de tipo numérico con las entradas del modelo para Torcoroma. y Fig. 33. Marco de datos de tipo numérico con las entradas del modelo para Belén. se puede observar los marcos de datos una vez convertidas todas las variables a tipo numérico de Ceiba, Torcoroma y Belén respectivamente.



	Hora	Minuto	Segundo	TemperaturaK	VelocidadDeViento
0	16	41	24	305.45	0.0
1	16	41	24	305.45	0.0
2	16	41	25	305.45	0.0
3	16	41	25	305.45	0.0
4	16	41	26	305.45	0.0
...
10166	18	6	11	302.55	1.0
10167	18	6	12	302.55	1.0
10168	18	6	12	302.55	1.0
10169	18	6	13	302.55	1.0
10170	18	6	13	302.55	1.0

10171 rows x 5 columns

Fig. 31. Marco de datos de tipo numérico con las entradas del modelo para Ceiba.

	Hora	Minuto	Segundo	TemperaturaK	VelocidadDeViento
0	15	23	32	311.35	0.0
1	15	23	32	311.35	0.0
2	15	23	33	311.35	0.0
3	15	23	33	311.35	0.0
4	15	23	34	311.35	0.0
...
10001	16	46	59	305.05	1.7
10002	16	47	0	305.05	1.9
10003	16	47	0	305.05	2.2
10004	16	47	1	305.05	2.2
10005	16	47	1	305.05	2.2

10006 rows x 5 columns

Fig. 32. Marco de datos de tipo numérico con las entradas del modelo para Torcoroma.

	Hora	Minuto	Segundo	TemperaturaK	VelocidadDeViento
0	21	19	47	298.85	1.8
1	21	19	48	298.85	1.8
2	21	19	48	298.85	1.6
3	21	19	49	298.85	1.6
4	21	19	49	298.85	1.6
...
7765	22	24	34	297.75	0.0
7766	22	24	35	297.75	0.0
7767	22	24	35	297.75	0.0
7768	22	24	36	297.75	0.0
7769	22	24	36	297.75	0.0

7770 rows x 5 columns

Fig. 33. Marco de datos de tipo numérico con las entradas del modelo para Belén.

C. Exploración de los datos

Para los datos experimentales los valores de velocidad de viento se encontraban en un rango muy aceptable, teniendo valores máximos no demasiados altos como para llegar a pensar que son

erróneos, sino que muy probablemente ocurrieron de forma natural. Además, estos pocos valores altos al no ser tan distantes no influirían demasiado en la precisión de los pronósticos. Es por esta razón que no fue necesario acotar los rangos de velocidad de viento. Por otro lado, tampoco se encontraron valores nulos o faltantes en ninguno de los marcos de datos correspondientes a las 3 zonas. Antes de agregar la variable de salida “Densidad de potencia” se analizó el perfil vertical del viento para las 3 zonas, de esta forma se puede observar en que zona hay más velocidad a una altura específica, lo cual se presenta en el apartado de selección de zona para mejor aprovechamiento del Recurso Eólico. Basado en la ecuación (8) y los valores diez minútales promedio de los datos experimentales se creó en Python los perfiles de viento para Ceiba, Torcoroma y Belén representados en las Fig. 34. Perfil vertical del viento para la zona 2 Ceiba., Fig. 35. Perfil vertical del viento para la zona 3 Torcoroma. y Fig. 36. Perfil vertical del viento para la zona 4 Belén. respectivamente.

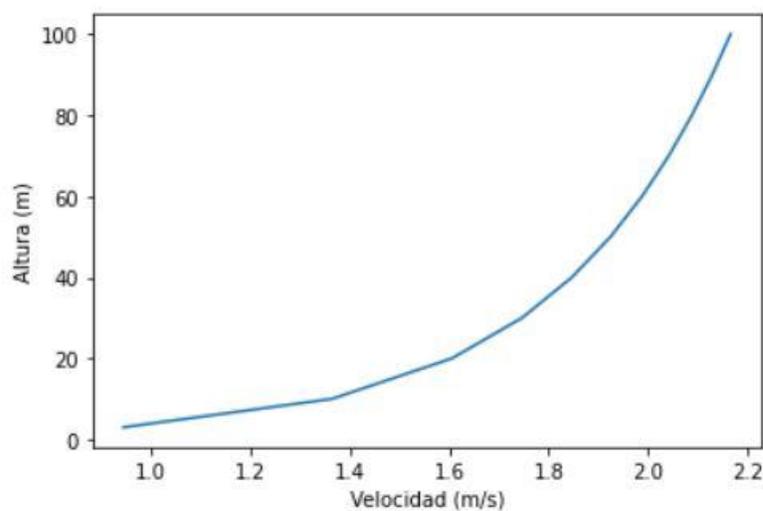


Fig. 34. Perfil vertical del viento para la zona 2 Ceiba.

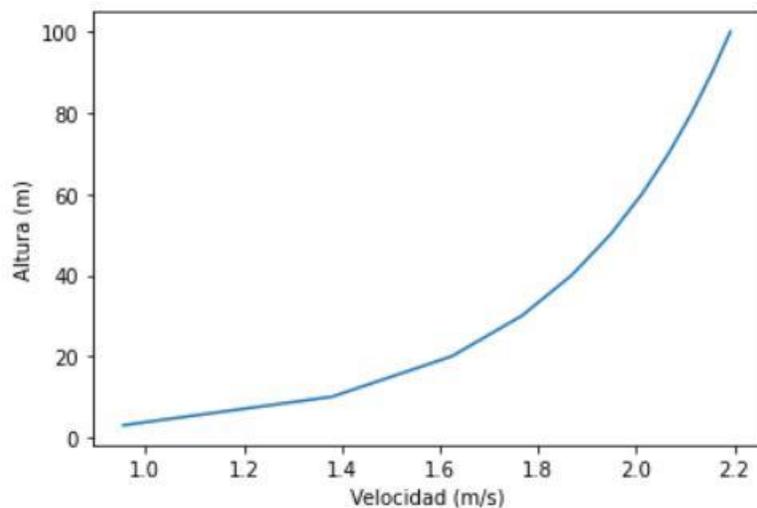


Fig. 35. Perfil vertical del viento para la zona 3 Torcoroma.

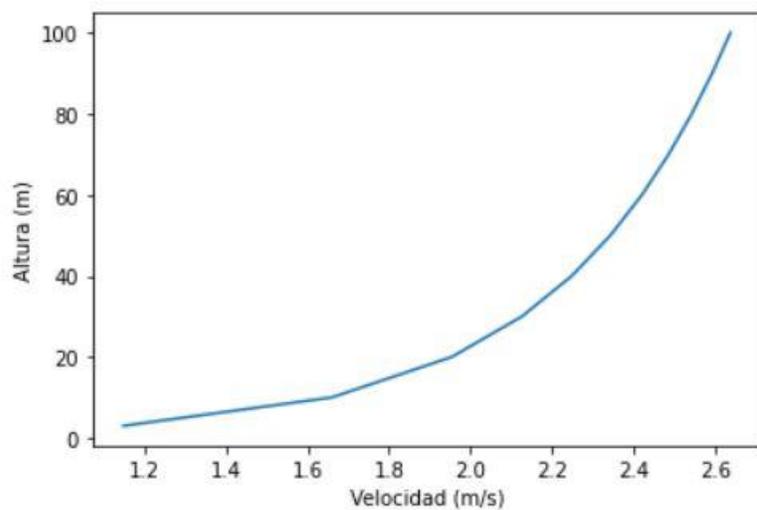


Fig. 36. Perfil vertical del viento para la zona 4 Belén.

Por último, se calculó la potencia disponible por área mediante la ecuación (6) con los parámetros constantes de presión a nivel del mar P_o , gravedad g , altura sobre el nivel del mar Z la constante universal sobre masa molecular del gas para el aire R_p , siguientes:

$$P_o = 101250 [Pa]$$

$$Z = 313 [m]$$

$$R_p = 287 \left[\frac{J}{kg \cdot K} \right]$$

$$g = 9.8 \left[\frac{m}{s} \right]$$

Los valores de Tk y v corresponden a los valores de temperatura y velocidad de viento del marco de datos experimental preprocesado del apartado anterior para cada zona. Los valores calculados de densidad de potencia fueron agregados al marco de datos con unidades de W/m^2 . En las Fig. 37. Resumen estadístico del marco de datos para Ceiba., Fig. 38. Resumen estadístico del marco de datos para Torcoroma. y Fig. 39. Resumen estadístico del marco de datos para Belén. puede observarse el resumen estadístico de todas las variables del marco de datos para cada zona.

	Hora	Minuto	Segundo	TemperaturaK	VelocidadDeViento	DensidadDePotencia
count	10171.000000	10171.000000	10171.000000	10171.000000	10171.000000	10171.000000
mean	16.854292	32.059679	29.505555	303.542026	0.952522	1.506079
std	0.521142	18.950421	17.324965	0.782404	0.752851	3.034726
min	16.000000	0.000000	0.000000	302.550000	0.000000	0.000000
25%	17.000000	14.000000	14.000000	302.850000	0.000000	0.000000
50%	17.000000	36.000000	30.000000	303.450000	1.000000	0.562070
75%	17.000000	49.000000	44.500000	304.150000	1.400000	1.544288
max	18.000000	59.000000	59.000000	305.450000	4.700000	58.411630

Fig. 37. Resumen estadístico del marco de datos para Ceiba.

	Hora	Minuto	Segundo	TemperaturaK	VelocidadDeViento	DensidadDePotencia
count	10006.000000	10006.000000	10006.000000	10006.000000	10006.000000	10006.000000
mean	15.563262	30.98331	29.581351	308.518319	0.975555	1.961814
std	0.496007	15.30126	17.333324	1.494712	0.872352	4.345188
min	15.000000	0.000000	0.000000	305.050000	0.000000	0.000000
25%	15.000000	20.000000	15.000000	307.750000	0.000000	0.000000
50%	16.000000	32.000000	30.000000	308.250000	1.000000	0.552915
75%	16.000000	43.000000	45.000000	308.750000	1.600000	2.259783
max	16.000000	59.000000	59.000000	312.450000	5.900000	113.485930

Fig. 38. Resumen estadístico del marco de datos para Torcoroma.

	Hora	Minuto	Segundo	TemperaturaK	VelocidadDeViento	DensidadDePotencia
count	7770.000000	7770.000000	7770.000000	7770.000000	7770.000000	7770.000000
mean	21.379665	28.919691	29.470013	298.035470	1.189691	2.872291
std	0.485335	16.791966	17.331865	0.166359	0.926501	5.194027
min	21.000000	0.000000	0.000000	297.750000	0.000000	0.000000
25%	21.000000	16.000000	14.000000	297.950000	0.000000	0.000000
50%	21.000000	27.000000	29.000000	297.950000	1.200000	0.986972
75%	22.000000	43.000000	44.750000	298.050000	1.800000	3.325650
max	22.000000	59.000000	59.000000	298.850000	4.800000	63.023455

Fig. 39. Resumen estadístico del marco de datos para Belén.

D. Entrenamiento del modelo

Una vez se tienen los marcos de datos experimentales en Python ya limpios en la etapa de preprocesamiento y exploración se puede proseguir con en el entrenamiento del modelo, para el cual se especifican a continuación sus parámetros.

E. Parámetros

Debido a que está usando el modelo ya evaluado anteriormente, los parámetros de número de épocas, número de capas y número de neuronas no cambian. Para el caso de las entradas en el modelo basado con datos experimentales se utilizaron las entradas Hora, Minuto, Segundo, Temperatura, Velocidad de viento, para un total de 5 entradas mientras que la salida corresponde a la variable Densidad de potencia, es decir, 1 salida. En este caso son 2 entradas menos que en el modelo anterior ya que las variables Año, Mes y Día no se utilizaron porque las mediciones se realizaron en un rango no mayor a un día. Con respecto a la partición de los datos se utilizó un 88% para entrenamiento y 12% para validación en las zonas de Ceiba y Torcoroma, mientras que para la zona de Belén se utilizó un 84% para entrenamiento y 16% para validación. La diferencia entre las proporciones de partición de los datos es debido a que no se contaban con la misma cantidad de datos para todas las zonas y se seleccionó de tal forma que se realizara predicciones de 10 minutos para cada una de ellas.

F. Evaluación del modelo

Para la evaluación del modelo con datos experimentales también se estimó por medio del comportamiento del error conforme aumentan las épocas, así como la proyección de valores reales promedio vs valores predichos promedio a 10 minutos. En la Tabla 27. Valores MAE para distintas épocas en cada zona. se muestran los valores de MAE conforme aumentan las épocas para cada zona. En las Fig. 40. Relación de Épocas vs MAE para zona 2 Ceiba., Fig. 41. Relación de Épocas vs MAE para zona 3 Torcoroma. y Fig. 42. Relación de épocas vs MAE para zona 4 Belén. se puede apreciar las gráficas de Épocas vs MAE para las zonas 2,3 y 4 respectivamente.

Tabla 27. Valores MAE para distintas épocas en cada zona.

Épocas	MAE Zona 2 Ceiba	MAE Zona 3 Torcoroma	MAE Zona 4 Belén
50	0.1093	0.6689	0.3903
100	0.0700	0.3532	0.3133
150	0.0413	0.1294	0.0385
200	0.0620	0.1720	0.2369
250	0.0266	0.0714	0.0713
300	0.0248	0.0331	0.0606

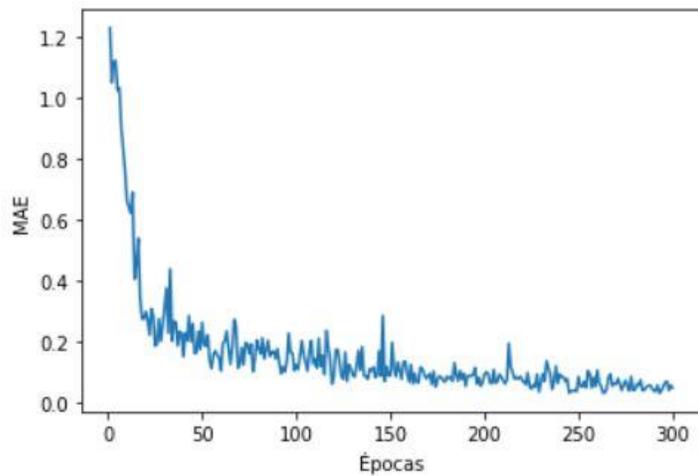


Fig. 40. Relación de Épocas vs MAE para zona 2 Ceiba.

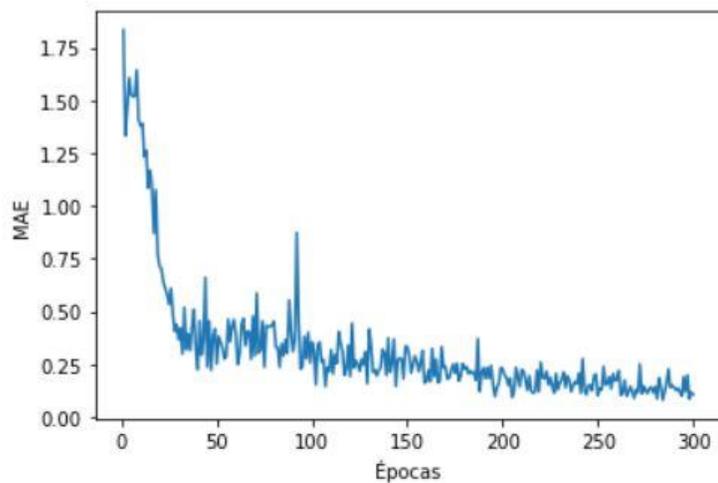


Fig. 41. Relación de Épocas vs MAE para zona 3 Torcoroma.

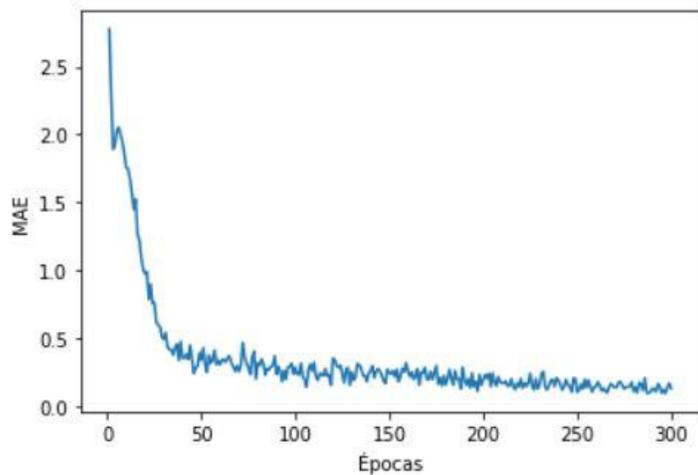


Fig. 42. Relación de épocas vs MAE para zona 4 Belén.

Para observar mejor la precisión del modelo se han presentado los valores promedio reales y valores promedio predichos en las Fig. 43. Valores reales y predichos para una predicción de 10 minutos en zona 2 Ceiba., Fig. 44. Valores reales y predichos para una predicción de 10 minutos en zona 3 Torcoroma. y Fig. 45. Valores reales y predichos para una predicción de 10 minutos en zona 4 Belén. para las 3 zonas.

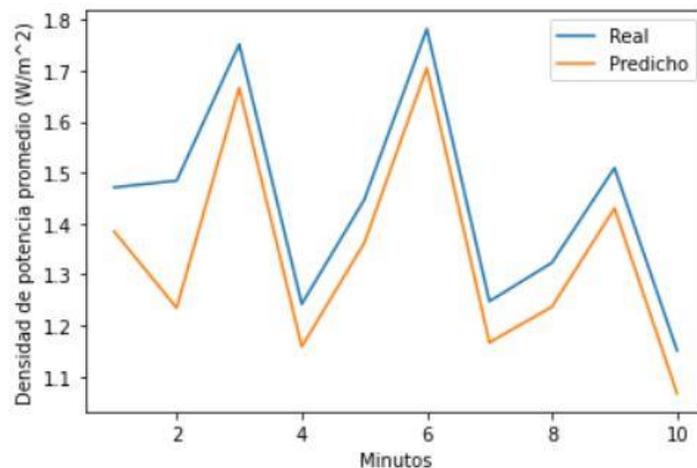


Fig. 43. Valores reales y predichos para una predicción de 10 minutos en zona 2 Ceiba.

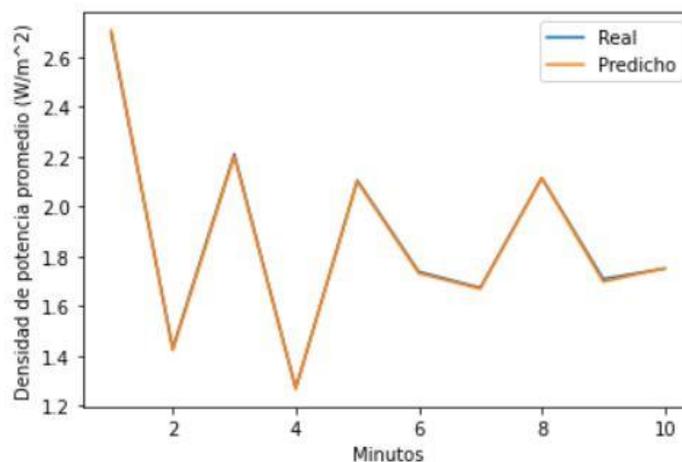


Fig. 44. Valores reales y predichos para una predicción de 10 minutos en zona 3 Torcoroma.

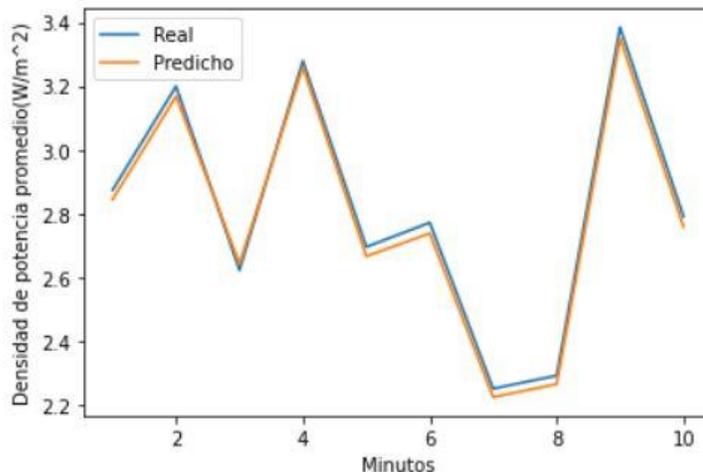


Fig. 45. Valores reales y predichos para una predicción de 10 minutos en zona 4 Belén.

9.3.9. Presentación de resultados mediante interfaz gráfica

En la Fig. 46. Interfaz gráfica con los resultados del modelo. se puede observar la presentación de los resultados obtenidos del modelo mediante la interfaz gráfica desarrollada en Python y en la Fig. 47. Resultados del modelo para la zona 1. se presentan los resultados que muestra la interfaz cuando se selecciona la zona 1 correspondiente al aeropuerto. En el siguiente apartado se observará las partes de la interfaz correspondientes a las zonas 2, 3 y 4.



Fig. 46. Interfaz gráfica con los resultados del modelo.

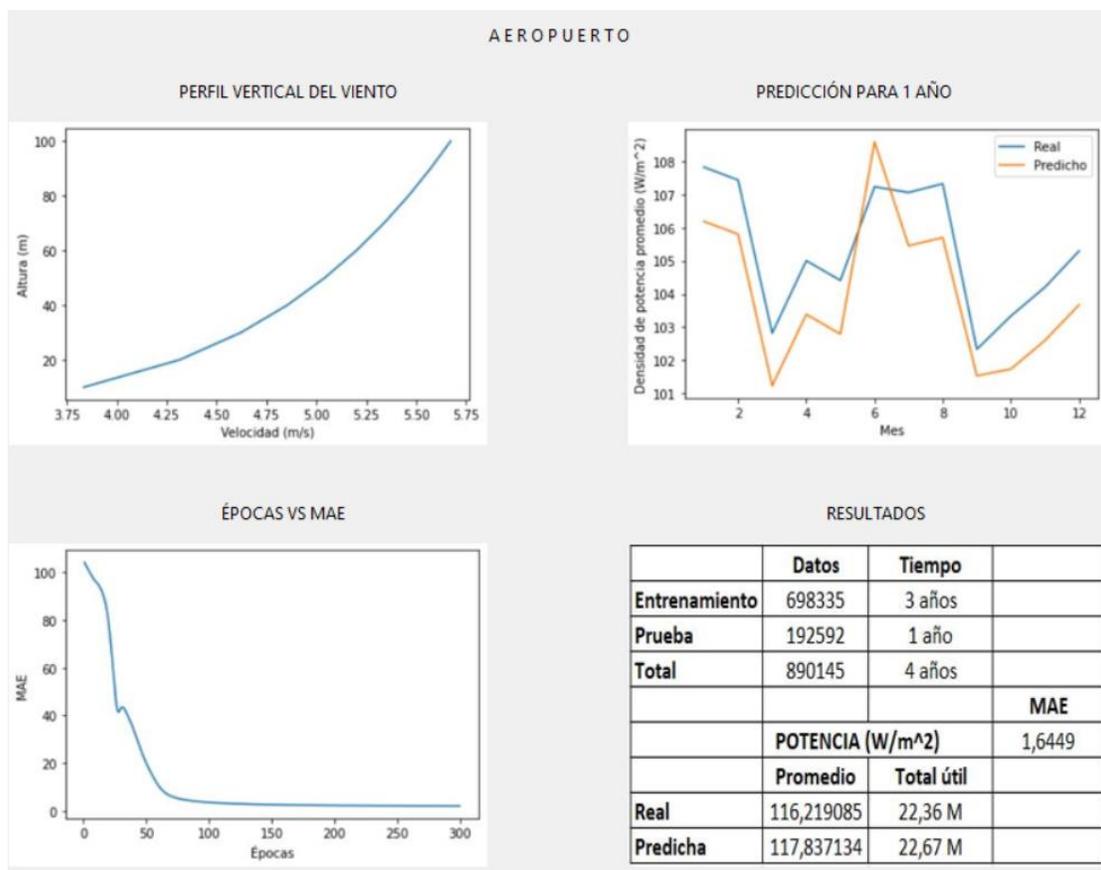


Fig. 47. Resultados del modelo para la zona 1.

9.4. Selección de zona para mejor aprovechamiento del Recurso Eólico

Con el fin de seleccionar la mejor zona para el aprovechamiento del Recurso Eólico es necesario analizar varios aspectos de los resultados obtenidos, dentro de los cuales se analizará una comparativa del perfil de viento de cada zona, así como el Recurso Eólico disponible.

9.4.1. Análisis del perfil vertical del viento

Es bien sabido que las aplicaciones para el aprovechamiento del Recurso Eólico se basan en Aerogeneradores o Mini aerogeneradores, los cuales funcionan a partir de la velocidad del viento y otros parámetros meteorológicos menos influyentes. Dependiendo de factores como el tipo de terreno, sus alrededores y los valores obtenidos tanto teóricos como experimentales se realizó la gráfica comparativa de la Fig. 48. Comparación del perfil vertical del viento de todas las zonas.

donde se puede apreciar que la zona de Aeropuerto alcanza mayores velocidades de viento conforme se aumenta su altura, esto es debido principalmente al ser una zona mucho más despejada que el resto. Al ser la densidad de potencia proporcional a la velocidad de viento, teóricamente esta zona puede ser capaz de generar más potencial eólico.

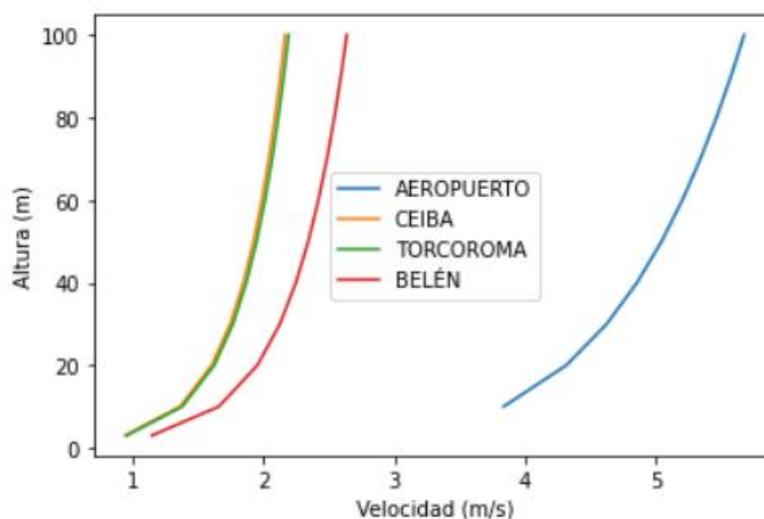


Fig. 48. Comparación del perfil vertical del viento de todas las zonas.

9.4.2. Análisis del Recurso Eólico disponible promedio y total

Es necesario recordar que las predicciones realizadas para la zona 1 del Aeropuerto fueron para un período de 1 año, mientras que para las zonas 2, 3 y 4 fueron para un período de apenas 10 minutos. En la Tabla 28. Recurso Eólico disponible promedio y total para la zona del Aeropuerto., Tabla 29. Recurso Eólico disponible promedio y total para la zona de Belén., Tabla 30. Recurso Eólico disponible promedio y total para la zona de Ceiba. y Tabla 31. Recurso Eólico disponible promedio y total para la zona de Torcoroma. se puede observar el análisis estadístico realizado para hallar la potencia eólica disponible promedio y total para las zonas que corresponden al Aeropuerto, Belén, Ceiba y Torcoroma respectivamente.

Tabla 28. Recurso Eólico disponible promedio y total para la zona del Aeropuerto.

VALORES PREDICHOS

Límite Inferior	Límite Superior	Frecuencia	Frecuencia Relativa	Centro Clase	Densidad*Fi
0	50	123594	0,642303675	25	16,05759187
50	100	18550	0,096402197	75	7,230164793
100	150	10932	0,056812335	125	7,101541915
150	200	7682	0,039922462	175	6,986430936
200	250	5096	0,026483321	225	5,958747135
250	300	4519	0,023484719	275	6,458297605
300	350	3522	0,018303425	325	5,948613212
350	400	2940	0,015278839	375	5,729564553
400	450	2677	0,013912058	425	5,91262479
450	500	1830	0,009510298	475	4,517391372
500	550	1685	0,008756749	525	4,597293463
550	600	1456	0,007566663	575	4,350831242
600	650	1261	0,006553271	625	4,095794162
650	700	997	0,005181293	675	3,497372975
700	750	872	0,004531683	725	3,285470032
750	800	675	0,003507897	775	2,718619915
800	850	687	0,003570259	825	2,9454639
850	900	464	0,002411354	875	2,109934883
900	950	431	0,002239857	925	2,071867708
950	1000	373	0,001938438	975	1,88997677
1000	1050	334	0,001735759	1025	1,77915322
1050	1100	303	0,001574656	1075	1,692755024
1100	1150	251	0,001304418	1125	1,467470105
1150	1200	125	0,00064961	1175	0,76329233
1200	1250	180	0,000935439	1225	1,145912911
1250	1300	202	0,001049771	1275	1,338457461
1300	1350	105	0,000545673	1325	0,723016479
1350	1400	121	0,000628823	1375	0,864631567
1400	1450	116	0,000602839	1425	0,859044917
1450	1500	76	0,000394963	1475	0,582570691
1500	1550	71	0,000368979	1525	0,56269261
1550	1600	73	0,000379373	1575	0,597511732
1600	1650	59	0,000306616	1625	0,498251249
1650	1700	37	0,000192285	1675	0,322076883
1700	1750	45	0,00023386	1725	0,403408116
1750	1800	33	0,000171497	1775	0,304407477
1800	1850	33	0,000171497	1825	0,312982336
1850	1900	16	8,31501E-05	1875	0,155906518
Densidad media de potencia eólica disponible W/m ²					117,8371349

Densidad de potencia eólica disponible W/m^2	22674575
--	----------

Tabla 29. Recurso Eólico disponible promedio y total para la zona de Belén.

VALORES PREDICHOS					
Límite Inferior	Límite Superior	Frecuencia	Frecuencia Relativa	Centro Clase	Densidad*Fi
0	2	823	0,649565904	1	0,649565904
2	4	231	0,182320442	3	0,546961326
4	6	35	0,027624309	5	0,138121547
6	8	77	0,060773481	7	0,425414365
8	10	23	0,018153118	9	0,163378058
10	12	16	0,012628256	11	0,138910813
12	14	12	0,009471192	13	0,123125493
14	16	12	0,009471192	15	0,142067877
16	18	12	0,009471192	17	0,16101026
18	20	6	0,004735596	19	0,089976322
20	22	2	0,001578532	21	0,033149171
22	24	6	0,004735596	23	0,108918706
24	26	0	0	25	0
26	28	1	0,000789266	27	0,021310182
28	30	1	0,000789266	29	0,022888713
30	32	1	0,000789266	31	0,024467245
32	34	4	0,003157064	33	0,10418311
34	36	1	0,000789266	35	0,027624309
36	38	0	0	37	0
38	40	1	0,000789266	39	0,030781373
40	42	0	0	41	0
42	44	0	0	43	0
44	46	0	0	45	0
46	48	0	0	47	0
48	50	1	0,000789266	49	0,038674033
50	52	2	0,001578532	51	0,08050513
Densidad media de potencia eólica disponible W/m^2					3,071033938
Densidad de potencia eólica disponible W/m^2					3891

Tabla 30. Recurso Eólico disponible promedio y total para la zona de Ceiba.

VALORES PREDICHOS					
Límite Inferior	Límite Superior	Frecuencia	Frecuencia Relativa	Centro Clase	Densidad*Fi

0	3	1074	0,87960688	1,5	1,319410319
3	6	85	0,06961507	4,5	0,313267813
6	9	39	0,031941032	7,5	0,23955774
9	12	9	0,007371007	10,5	0,077395577
12	15	6	0,004914005	13,5	0,066339066
15	18	4	0,003276003	16,5	0,054054054
18	21	1	0,000819001	19,5	0,015970516
21	24	1	0,000819001	22,5	0,018427518
24	27	0	0	25,5	0
27	30	0	0	28,5	0
30	33	0	0	31,5	0
33	36	1	0,000819001	34,5	0,028255528
36	39	0	0	37,5	0
39	42	1	0,000819001	40,5	0,033169533
Densidad media de potencia eólica disponible W/m^2					2,165847666
Densidad de potencia eólica disponible W/m^2					2644,5

Tabla 31. Recurso Eólico disponible promedio y total para la zona de Torcoroma.

VALORES PREDICHOS					
Límite Inferior	Límite Superior	Frecuencia	Frecuencia Relativa	Centro Clase	Densidad*Fi
0	2,5	931	0,775187344	1,25	0,96898418
2,5	5	131	0,10907577	3,75	0,409034138
5	7,5	75	0,06244796	6,25	0,39029975
7,5	10	31	0,025811823	8,75	0,225853455
10	12,5	8	0,006661116	11,25	0,074937552
12,5	15	8	0,006661116	13,75	0,091590341
15	17,5	4	0,003330558	16,25	0,054121565
17,5	20	4	0,003330558	18,75	0,06244796
20	22,5	2	0,001665279	21,25	0,035387177
22,5	25	0	0	23,75	0
25	27,5	3	0,002497918	26,25	0,065570358
27,5	30	0	0	28,75	0
30	32,5	1	0,000832639	31,25	0,026019983
32,5	35	0	0	33,75	0
35	37,5	1	0,000832639	36,25	0,030183181
37,5	40	1	0,000832639	38,75	0,032264779
40	42,5	0	0	41,25	0
42,5	45	0	0	43,75	0
45	47,5	0	0	46,25	0

47,5	50	0	0	48,75	0
50	52,5	0	0	51,25	0
52,5	55	0	0	53,75	0
55	57,5	0	0	56,25	0
57,5	60	0	0	58,75	0
60	62,5	0	0	61,25	0
62,5	65	0	0	63,75	0
65	67,5	0	0	66,25	0
67,5	70	0	0	68,75	0
70	72,5	0	0	71,25	0
72,5	75	0	0	73,75	0
75	77,5	0	0	76,25	0
77,5	80	0	0	78,75	0
80	82,5	1	0,000832639	81,25	0,067651957
Densidad media de potencia eólica disponible W/m^2					2,534346378
Densidad de potencia eólica disponible W/m^2					3043,75

Luego de obtener las predicciones del Recurso Eólico disponible promedio y total, además del perfil vertical del viento de todas las zonas, se comprobó la relación existente entre la velocidad y la generación de energía eólica, ya que las zonas en donde se alcanzaban mayores velocidades de viento conforme aumentaban la altura, fueron las zonas que más Recurso Eólico disponible predijeron, lo que confirma la relación proporcional entre velocidad y potencial eólico. De las 3 zonas escogidas para el estudio, Belén fue la que registró una mayor predicción del Recurso Eólico a corto plazo, convirtiéndose en una opción más viable para el aprovechamiento del mismo.

10. Conclusiones

El estudio del Recurso Eólico abarca una cantidad de factores a tener en cuenta, tanto teóricamente como a nivel experimental, en donde se busca obtener la cantidad de energía que puede extraerse del desplazamiento del viento. Para ello se tiene en cuenta en primera instancia el Recurso Eólico que el viento en una zona específica puede generar en función de la altura, lo que se conoce como el Recurso Eólico disponible o también llamada útil. Por otra parte, con el fin de aprovechar esta energía por medio de aerogeneradores se estudia cuánta energía realmente se puede aprovechar del Recurso Eólico disponible, lo cual queda fuera de los límites de este proyecto.

La literatura consultada sobre el Recurso Eólico permitió recopilar en conjunto de información fundamental para la selección de la técnica y herramienta de software, que mejor se adaptara a las necesidades del proyecto.

El ajuste de la partición de los datos a 78.4% para entrenamiento y 21.6% para validación permitió un pronostico exacto para 1 año, con un total de 890145 datos usados para la construcción del modelo se alcanzó un MAE de 1.6, lo cual representa el 1.7% del valor promedio.

Para el uso del modelo se alcanzó un MAE del 0.02, 0.03 y 0.06 para las zonas de Ceiba, Torcoroma y Belén respectivamente. Siento Belén la que registro un mayor perfil vertical del viento y un mayor Recurso Eólico disponible a corto plazo, donde se verificó para cada una de las zonas que la velocidad del viento y la energía eólica son directamente proporcionales.

Aumentar el número de épocas y la cantidad de datos aumentaron considerablemente la precisión del modelo tanto para el entrenamiento como para la parte experimental. Si bien puede seguirse aumentando el numero de épocas para que sea aún más preciso el modelo, no es recomendable ya que la relación entre numero de épocas y MAE no es de forma lineal sino de

forma logarítmica, además, de que al seguir aumentando más recursos computacionales serán necesarios y los tiempos de simulación serán mas largos.

11. Recomendaciones

Para futuros trabajos relacionados donde sea necesario la toma de datos experimentales de la velocidad del viento y/o temperatura, se recomienda realizar las mediciones siempre a una misma altura con respecto al suelo, con el fin de ahorrar cálculos de extrapolación mediante perfil vertical de viento y aprovechar mayor cantidad de datos en las mismas condiciones. Ya sea que se utilice un anemómetro u otro tipo de sensor, se recomienda fijar éste a una base donde fuertes vientos no afecten su posición ni sus mediciones. Con respecto a los horarios de medición es importante que sean los más variados posibles, por ejemplo, si el tiempo total de medición va a estar en el rango de los días, se recomienda realizar mediciones a distintas horas del mismo, si el rango está en meses, tomar mediciones distintos días de las semanas y a distintas horas, si está en el rango del año, tomar mediciones a lo largo de los meses, en distintos días y horas. Esta recomendación se base en que mientras más datos y mejor distribución de horarios en las mediciones, se puede obtener información más cercana a la realidad.

Con respecto a la evaluación del modelo de la Red Neuronal Artificial se recomienda realizar varias veces el entrenamiento para un mismo número de épocas y luego sacar la media de los resultados obtenidos, ya que debido a que la Red trabaja con pesos aleatorios para las neuronas, va a arrojar diferentes resultados cada vez que se ejecute. La idea es verificar que estos resultados no difieran mucho y presentar una métrica de éxito promedio que refleje la precisión del modelo.

Como trabajo futuro queda el estudio de cuanta energía realmente se puede aprovechar del Recurso Eólico disponible respetando factores como por ejemplo el límite de Betz, el cual establece que la máxima potencia eólica aprovechable es de un 59% de la disponible. Un estudio más profundo podría abarcar este tema a cabalidad junto con casos de aplicación donde se evalúe el rendimiento de aerogeneradores o miniaerogeneradores teniendo en cuenta las características

del fabricante y la energía aprovechable.

Referencias

- Aguilar, R. M., Torres, J. M., & Martín, C. A. (2019). Automatic learning for the system identification. A case study in the prediction of power generation in a wind farm. *RIAI - Revista Iberoamericana de Automatica e Informatica Industrial*, 16(1), 114–127. <https://doi.org/10.4995/riai.2018.9421>
- Alberto Ruiz Marta Susana Basualdo Autor, C., & Jorge Matich, D. (n.d.). *Cátedra: Informática Aplicada a la Ingeniería de Procesos-Orientación I Redes Neuronales: Conceptos Básicos y Aplicaciones*.
- April Speight. (2021). *Visual Studio Code for Python Programmers* (Wiley, Ed.).
- Autónoma de Occidente Colombia Vera Duarte, U., Vega, V., & Serrano, R. (n.d.). *El Hombre y la Máquina*. <http://www.redalyc.org/articulo.oa?id=47843368017>
- Badii, M., Guillen, A., & Abreu, J. (2016). Energías Renovables y Conservación de Energía. *Revista Daena (International Journal of Good Conscience)*, 11(1), 141–155. <http://bd.univalle.edu.co/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=fua&AN=116281297&lang=es&site=eds-live>
- Camacho, Y. (n.d.). *APROVECHAMIENTO DEL RECURSO EÓLICO Y SOLAR EN LA GENERACIÓN DE ENERGÍA ELÉCTRICA Y LA REDUCCIÓN DE EMISIONES DE CO2 EN EL POBLADO RURAL LA GRAMITA DE CASMA*.
- Cámara de Comercio de Cali. (2016, April 13). *INFORMES ECONÓMICOS*. Informe #03.
- de la Cruz Buelvas, J., Valencia Ochoa, G., & Vanegas Chamorro, M. (2018). Estudio estadístico de la velocidad y la dirección del viento en los departamentos de Atlántico y Bolívar en Colombia. *Ingeniare. Revista Chilena de Ingeniería*, 26(2), 319–328. <https://doi.org/10.4067/s0718-33052018000200319>

- Delgado, D. A. P., Herrera, M. T. J., & Almeida, J. E. (2014). La energía eólica como energía alternativa para el futuro de Colombia. *El Centauro*, 6(9), 111–120.
- Deusto. Biblioteca Universitaria. Biblioguías LibGidak. (n.d.).
- Encinas, E. Torrez, R. Ghezzi, F. Palenque, E. (2015). *CARACTERÍSTICAS DEL VIENTO Y DEL POTENCIAL EÓLICO DEL ALTIPLANO CENTRAL*. 20–24.
- Faustino Menéndez Díez. (2006). *Higiene Industrial* (L. Nova, Ed.).
- Fe, S. (2010). *Caracterización de la variabilidad del suelo*. 27–30.
- Fernández Estrella, S. K., Olivares Gutiérrez, L. M., & Zamora Rasgado, F. J. (2015). Inteligencia Artificial. *XIKUA Boletín Científico de La Escuela Superior de Tlahuelilpan*, 3(5). <https://doi.org/10.29057/xikua.v3i5.1271>
- García, U. A., Ibarguengoytia, P. H., Díaz González, L., & Hermosillo Valadez, J. (2019). Técnicas de aprendizaje automático en el diagnóstico de aerogeneradores. *Revista de Energías Renovables*, 3(11), 7–14. <https://doi.org/10.35429/jre.2019.11.3.7.14>
- Giraldo, M., Vacca Ramírez, R., & Urrego Quintanilla, A. (2018). Las Energías Alternativas ¿Una Oportunidad Para Colombia? *Punto de Vista*, 9(13). <https://doi.org/10.15765/pdv.v9i13.1117>
- González, M. (2009). La generación eléctrica a partir de combustibles fósiles. *Divulgacion*, 2, 143–151.
- Google Research. (n.d.).
- Guerrero Hoyos, B. G., Vélez Macías, F. D. J., & Morales Quintero, D. E. (2020). Energía eólica y territorio: sistemas de información geográfica y métodos de decisión multicriterio en La Guajira (Colombia). *Ambiente y Desarrollo*, 23(44). <https://doi.org/10.11144/javeriana.ayd23-44.eets>

- Henao león, D., Báez Alarcón, A. C., & Pedroza Rojas, J. B. (2018). Metodología para determinar la viabilidad de generación de energía eléctrica por medio del recurso eólico. *Investigacion e Innovación En Ingenierias*, 6(2), 6–15. <https://doi.org/10.17081/invinno.6.2.3108>
- I, P. F. B. (2020). *Neural networks in predicting wind speed Redes neurais na previsão da velocidade do vento* (Vol. 5, Issue 05). <https://doi.org/10.23857/pc.v5i5.1398>
- IDEAM. (n.d.). *Consulta y Descarga de Datos Hidrometeorológicos*.
- Invierta y Gane con Energía Guía práctica para la aplicación de los incentivos tributarios de la Ley 1715 de 2014*. (n.d.).
- Larranaga, P. (n.d.-a). *Tema 8. Redes Neuronales*. <https://www.researchgate.net/publication/268291232>
- Larranaga, P. (n.d.-b). *Tema 8. Redes Neuronales The correlation between the brain activity and energy metabolism View project Regularized Model learning in EDAs for continuous and multi-objective optimization View project*. <https://www.researchgate.net/publication/268291232>
- Ley_143_1994 (1)*. (n.d.).
- Londoño, A. (2017). *Energías renovables en Colombia: una aproximación desde la economía Renewable Energy in Colombia: An Approach from the Economy*.
- Miguel Villarrubia López. (2012). *Ingeniería de la Energía Eólica*. MARCOMBO, S.A.
- Miranda, R. M. (n.d.). *Redes Neuronales Artificiales*.
- Posso, F., Acevedo Paez, J. C., & Hernández, J. (2014). El impacto económico de las energías renovables. *Aibi Revista de Investigación, Administración e Ingeniería*, July, 22–26. <https://doi.org/10.15649/2346030x.526>

Predicción de Potencia eólica utilizando técnicas modernas de inteligencia artificial. (n.d.).

<https://doi.org/10.22201/fi.25940732e.2018.19n4.033>

Ramírez, S., Guamán Molinan, J., Ríos Villacorta, A., Guevara, D., & Nogales, D. (2019).

Plataforma Web de predicción de viento basado en técnicas de inteligencia artificial con el uso del método ARIMA: Caso de estudio - ProQuest. *Risti*, 404–417.

<https://search.proquest.com/openview/72add25354599e3f8115f573512f071d/1?pq-origsite=gscholar&cbl=1006393>

Rojas, E. M. (n.d.). *Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo.*

Rouhiainen, L. (n.d.). *Inteligencia artificial : 101 cosas que debes saber hoy sobre nuestro futuro.*

Sandoval Serrano, L. (2018). *Algoritmos de aprendizaje automático para análisis y predicción de datos.* 36–40.

Santamarta, J. (2007). NP 4456: Sistema de Gestão de IDI. *Ipq.*

[http://onlinebiblio.ineg.pt/plinkres.asp?Base=ENERBD&Form=COMP&StartRec=0&RecPag=5&NewSearch=1&SearchTxt=%22TCO NP 4456 \(2007\) - norma portuguesa : gest?o da investiga??o, desenvolvimento, e inova??o \(IDI\): terminologia e defini??es](http://onlinebiblio.ineg.pt/plinkres.asp?Base=ENERBD&Form=COMP&StartRec=0&RecPag=5&NewSearch=1&SearchTxt=%22TCO NP 4456 (2007) - norma portuguesa : gest?o da investiga??o, desenvolvimento, e inova??o (IDI): terminologia e defini??es)

Secretaría Jurídica Distrital. (2003, December 19). *Decreto 3683 de 2003 Nivel Nacional* .
Diario Oficial 45409 de Diciembre 22 de 2003.

Tripp-Barba, C. (2020). Revisión de literatura de modelos computacionales para la predicción de la velocidad del viento de 2004 a 2016. *Revista de Investigación En Tecnologías de La Información*, 8(15), 27–40. <https://doi.org/10.36825/riti.08.15.004>

Universitat de Valencia. (2018). *Tecnologías de la información geográfica: Perspectivas*

multidisciplinares en la sociedad del conocimiento (Publicacions de la Universitat de Valencia, Ed.).

Vidal, C., Bravo, J., Cajiao, E., Meza, P., & Arango, S. (n.d.). *Guía metodológica para la priorización de proyectos: Un enfoque aplicado a la infraestructura, la logística y la conectividad*.

Zhang, X.-D. (2020). *Machine Learning BT - A Matrix Algebra Approach to Artificial Intelligence* (X.-D. Zhang, Ed.; pp. 223–440). Springer Singapore.
https://doi.org/10.1007/978-981-15-2770-8_6

Anexos

Anexo 1. Ponencia: Análisis del Recurso Eólico disponible en la ciudad de Cúcuta usando aprendizaje automático

G. F. Di Candia, B. O. Calderón, S. A. Castro, B. Medina and D. Guevara. “Análisis del Recurso Eólico disponible en la ciudad de Cúcuta usando aprendizaje automático” ponencia aceptada en el VI Congreso de Investigación, Desarrollo e Innovación IDI-UNICyT-2021.



BRAYAN ORLANDO CALDERON ANGARITA <brayanorlandoca@ufps.edu.co>

Fwd: Notificación aceptación trabajo para el IDI-UNICYT 2021 Gian Franco Di Candia Jaimes

1 mensaje

Gian Franco Di Candia Jaimes <jaimesgianfrancodc@ufps.edu.co>

29 de noviembre de 2021, 10:38

Para: Sergio Castro Casadiego <sergio.castroc@ufps.edu.co>, brayanorlandoca@ufps.edu.co

----- Forwarded message -----

From: **Contacto eSimposio** <postmaster@esimposio.com>

Date: Mon, Nov 29, 2021, 9:33 AM

Subject: Notificación aceptación trabajo para el IDI-UNICYT 2021 Gian Franco Di Candia Jaimes

To: Gian Franco Di Candia Jaimes <jaimesgianfrancodc@ufps.edu.co>, Control 1 <investigaciones@unicyt.net>



**GRACIAS POR FORMAR
PARTE DE NUESTRO CONGRESO**

Estimado/a GIAN FRANCO DI CANDIA JAIMES,

En nombre del Comité Organizador tenemos el agrado de informarle que ha concluido el proceso final de revisión del resumen del trabajo titulado **ANÁLISIS DEL RECURO EÓLICO DISPONIBLE EN LA CIUDAD DE CÚCUTA USANDO APRENDIZAJE AUTOMÁTICO - Código IDIUNICYT032** del cual usted es el/la autor/a, y este ha sido **ACEPTADO** para ser presentado en el VI Congreso de Investigación, Desarrollo e Innovación (IDI-UNICYT 2021).

El congreso usará la plataforma streamyard.com para la transmisión. Le invitamos a participar en un taller de capacitación que se dictará el miércoles 1 de diciembre de 5:00 a 6:00 p.m.

El enlace para conectarse al Zoom es:

<https://us06web.zoom.us/j/86417354314?pwd=OENKOUhBTWtjV1RDNIzsREdKaDQzUT09>

ID de reunión: 864 1735 4314
Código de acceso: 594667

Anexo a esta comunicación encontrará información importante sobre su participación en el evento. Cualquier información adicional puede solicitarla escribiendo al correo investigaciones@unicyt.net

Agradecemos su atención,

Comité Organizador IDI-UNICYT 2021

Bloque A1: Ciencia y Tecnología

13/6/22, 2:08

Correo de Universidad Francisco de Paula Santander - Fwd: Notificación aceptación trabajo para el IDI-UNICYT 2021 Gian Franco Di Candia Jaimes
Su trabajo ha sido aprobado en la **MODALIDAD ORAL**

Tiempo de presentación: 15 minutos.

Día: Viernes 3 de diciembre

Hora: Entre 1:30 p.m. y 5:00 p.m.



Evento gestionado con:



Click aqui para no recibir más correos / Unsubscribe

[unsubscribe](#)

Análisis del Recurso Eólico disponible en la ciudad de Cúcuta usando Aprendizaje Automático

Analysis of available Wind Resource in city of Cúcuta using Machine Learning

G. F. Di Candia¹, B.O. Calderón², S.A. Castro³, B. Medina⁴, D. Guevara⁵

¹Universidad Francisco de Paula Santander (UFPS),

²Universidad Francisco de Paula Santander (UFPS),

³Universidad Francisco de Paula Santander (UFPS),

⁴Universidad Francisco de Paula Santander (UFPS),

⁵Universidad Francisco de Paula Santander (UFPS),

¹jaimessgianfrancodc@ufps.edu.co – 0000-0002-8517-7790,

²brayanorlandoca@ufps.edu.co – 0000-0001-7551-2885,

³sergio.castroc@ufps.edu.co – 0000-0003-0962-9916,

⁴byronmedina@ufps.edu.co – 0000-0003-0754-8629,

⁵dinaelgi@ufps.edu.co – 0000-0003-3007-8354

Resumen

Hoy en día el uso de las energías renovables está creciendo en todo el mundo debido a que se trata de energías limpias que no impactan negativamente al medio ambiente, no afecta la salud de las personas, y porque contribuye a mejorar la eficiencia energética. Dentro de las energías renovables más usadas se encuentra la energía del viento, la cual se deriva principalmente del movimiento de las masas de aire. Esta energía depende de varios parámetros meteorológicos como lo son velocidad de viento, temperatura, dirección del viento, humedad, presión atmosférica, altura y otros. Para el presente proyecto se escogió las dos variables más influyentes como velocidad de viento y temperatura, las cuales fueron tomadas experimentalmente en 3 sectores de la ciudad Cúcuta, con el fin de crear un modelo que prediga el Recurso Eólico disponible a corto plazo en cada una de las zonas seleccionadas. Para la construcción del modelo se utilizó el enfoque del Aprendizaje Automático en conjunto con la técnica de Redes Neuronales Artificiales. La Red fue

entrenada mediante el algoritmo de Retropropagación (del inglés *backpropagation*) junto con validación cruzada (del inglés *cross-validation*) para evaluar los resultados del modelo, garantizando la independencia entre datos de entrenamiento y validación. Se utilizaron 300 épocas para entrenar el modelo y las predicciones obtenidas para un rango de 10 minutos presentaron un Error Medio Absoluto (MAE) de 0.02, 0.03 y 0.06 en las zonas de Ceiba, Torcoroma y Belén respectivamente.

Palabras clave: Aprendizaje Automático, Error Medio Absoluto, Redes Neuronales Artificiales, Retropropagación, Validación cruzada.

Abstract

Currently the use of renewable energies is growing around the world because they are clean energies that do not negatively impact the environment, do not affect people's health, and because they contribute to improving energy efficiency. Among the most used renewable energies is wind energy, which is derived mainly from the movement of air masses. This energy depends on various meteorological parameters such as wind speed, temperature, wind direction, humidity, atmospheric pressure, height and others. For the present project, the two most influential variables such as wind speed and temperature were chosen, which were taken experimentally in 3 sectors of the city of Cúcuta, in order to create a model that predicts the Wind Resource available in the short term in each one of the selected zones. For the construction of the model, the Machine Learning approach was used in conjunction with the Artificial Neural Networks technique. The Network was trained using the Backpropagation algorithm and Cross-validation to evaluate the results of the model, guaranteeing the independence between training and validation data. 300 epochs were used to train the model and the predictions obtained for a range of 10 minutes presented a Mean Absolute Error (MAE) of 0.02, 0.03 and 0.06 in the areas of Ceiba, Torcoroma and Belén respectively.

Keywords: Artificial Neural Networks, Backpropagation, Cross-validation, Machine Learning, Mean Absolute Error.

1. Introducción

Es bien conocido que los combustibles fósiles como el carbón o el petróleo han sido utilizados

durante un gran periodo de tiempo para generar energía. Este proceso ha traído consigo una serie de eventos que cada día son más causantes de preocupación a nivel mundial en temas como: el impacto negativo al medio ambiente producido por los gases de efecto invernadero (GEI) y los efectos sobre la salud de las personas. Es por esta razón que cada vez son más los países que se suman a la necesidad de fomentar el desarrollo de energías renovables (Guerrero Hoyos et al., 2020).

Las energías renovables son un tipo de energía que se genera a partir de recursos naturales como el viento, el agua y la luz solar (Giraldo et al., 2018). El aprovechamiento de las energías renovables en todo el mundo se está convirtiendo en el método más eficaz y consciente con el medio ambiente para la generación de energía eléctrica. Dentro de estas energías también llamadas energías limpias, se encuentra la energía eólica la cual se deriva del movimiento de las masas de aire, es decir, la velocidad del viento. Sin embargo, otros parámetros como la temperatura pueden ser relevantes para estimar la energía eólica. Es bien sabido que los datos meteorológicos suelen ser proporcionados por estaciones meteorológicas ubicadas estratégicamente en algunos puntos de la ciudad, lo cual deja una gran cantidad de zonas descubiertas donde no es posible saber con certeza el valor exacto de la energía eólica disponible. Es por esta razón que se ha propuesto realizar un modelo de predicción del recurso eólico en distintas zonas de la ciudad mediante la toma de datos experimentales. Esto permitiría tener un acercamiento más real al potencial eólico disponible en las zonas seleccionadas para el estudio.

2. Materiales y métodos

2.1 Técnicas de Inteligencia Artificial

Dentro de los enfoques de la Inteligencia Artificial se encuentra el Aprendizaje Automático, definida como la ciencia capaz de enseñar a las máquinas a realizar predicciones por sí mismas, este aprendizaje se realiza alimentando el algoritmo con datos. Es importante tener en cuenta que las primeras predicciones que suele arrojar estos modelos son imprecisas, pero a medida que se va aumentando el número de predicciones la computadora actualizará su algoritmo y mejorando de esta forma los pronósticos (Norman, 2021). Para llevar a cabo el enfoque del Aprendizaje Automático se utilizan diversas técnicas de Inteligencia Artificial como Redes Neuronales Artificiales, Redes Bayesianas, Máquinas de vectores de soporte, entre otras. A pesar de la gran

diversidad de técnicas utilizadas en el campo de las predicciones, las Redes Neuronales Artificiales están representando actualmente mayor diversidad de campos y aplicaciones debido a su habilidad para calcular funciones de salida no lineales y porque sus cotas de error son mucho mejores que las proporcionadas por métodos estadísticos. Además, normalmente no tiene restricciones con las variables de entradas, lo que quiere decir que no importa si los datos de entrada tienen alguna relación funcional o no, lo que hace a esta técnica ideal para la resolución de problemas complejos (Palmer Pol & Montaña Moreno, s.f.). Por todas estas razones se ha seleccionado la técnica de Redes Neuronales Artificiales para realizar el entrenamiento en el Aprendizaje Automático.

2.2 Lenguaje de programación

En vista de que hoy en día son muchos los lenguajes de programación empleados en el uso del Aprendizaje Automático, se reunió la información más relevante de cada uno de ellos para estudiar sus características, ventajas y desventajas. Los lenguajes comparados fueron Python, R, Julia y Matlab. El lenguaje seleccionado fue Python por tener la mayor comunidad y recursos en internet, su simpleza que lo hace ideal para trabajar grandes proyectos en pocas manos producto de la gran cantidad y diversidad de sus bibliotecas, lo que lo hace una opción muy eficiente para ahorrar tiempo en el desarrollo (Rojas, 2020.).

2.3 Entorno de programación

El entorno utilizado para realizar el modelo fue Google Colaboratory, el cual es un producto de Google Research y permite desarrollar código en Python en el navegador. Es útil para desarrollar aplicaciones de aprendizaje automático, análisis de datos y educación, y es que Google Colaboratory permite la entrada de texto lo que hace más dinámico su uso. Este entorno no requiere instalaciones de paquetes ni configuraciones. Además, cuenta con una GPU y otros recursos computacionales gratuitos (Google Research, s.f.). La ventaja de contar con una GPU permitió acelerar significativamente los procesos de entrenamiento del algoritmo de la Red Neuronal Artificial, lo cual fue de suma importancia teniendo en cuenta que se utilizaron alrededor de 10 mil datos por cada zona.

2.4 Bases de datos y ecuaciones

Los datos de velocidad de viento y temperatura fueron tomados con un anemómetro de álabe en

las zonas de Ceiba, Torcoroma y Belén. En la zona de Ceiba se tomaron 10171 datos, en Torcoroma 10006 y en Belén 7770. En la Figura 1 puede observarse el anemómetro con que se tomaron los datos.



Figura 1. Anemómetro fijado en la zona de Torcoroma. Elaboración y fuente propias

El principal factor para la generación de energía eólica es el viento, el cual es generado por el movimiento de la masa atmosférica. Sin embargo para la estimación del potencial eólico es necesario conocer no sólo la velocidad del viento sino también otros parámetros como la dirección del viento, densidad y humedad atmosférica (Villarrubia, 2012).

A partir de los datos obtenidos de velocidad de viento y temperatura, se procedió a utilizar la ecuación (1) para calcular la densidad de potencia para cada uno de los valores obtenidos.

La forma en que se suele estimar el Recurso Eólico es mediante la potencia que se puede extraer del viento. Si se tiene una masa de aire m con velocidad v , la energía cinética generada estará dada por la ecuación 1.

$$E = \frac{1}{2}mv^3 \quad (1)$$

El caudal de masa de aire m de densidad ρ que fluye con velocidad v por medio de una superficie de área A ortogonal a la dirección del flujo se puede observar en la ecuación 2.

$$Pd = \frac{1}{2}mv^3 = \frac{1}{2}\rho Av^3 \quad (2)$$

La densidad de potencia disponible o potencia disponible por unidad de área (Pd/A) se puede apreciar en la ecuación 3.

$$\frac{Pd}{A} = \frac{1}{2}\rho v^3 \quad (3)$$

Donde Pd es la potencia disponible, A es el área, ρ es la densidad del aire seco y v es la velocidad del viento en [m/s]. Para calcular la densidad del aire seco se utiliza la siguiente expresión:

$$\rho = \frac{Patm}{Rp.Tk} \quad (4)$$

Donde $Patm$ es la presión atmosférica, Rp es la constante universal sobre masa molecular del gas para el aire y Tk es la temperatura en Kelvin. Por otro lado, la presión atmosférica $Patm$ se calcula mediante:

$$Patm = Po.e^{-\left(\frac{g.Z}{Rp.Tk}\right)} \quad (5)$$

Donde Po es la presión a nivel del mar es [Pa], g es la gravedad en [m/s^2], Z es la altura sobre el nivel del mar de la zona en [m]. Por último la constante Rp y la variable Tk definida anteriormente como la constante universal sobre masa molecular del gas para el aire y la temperatura en Kelvin respectivamente.

Si se sustituye la ecuación (5) en la ecuación (4) y luego en la ecuación (3) se obtiene la ecuación (6).

$$\frac{Pd}{A} = \frac{Po.e^{-\left(\frac{g.Z}{Rp.Tk}\right)}.v^3}{2.Rp.Tk} \left[\frac{W}{m^2} \right] \quad (6)$$

Con la ecuación (6) se obtuvieron los valores de densidad de potencia para cada par temperatura Tk y velocidad v .

Por otro lado, la velocidad del viento medio o estacionario se obtiene a partir del promedio de velocidad para intervalos de tiempo de 10 minutos (Villarrubia, 2012). Su variación con respecto

a la altura se corresponde con la expresión:

$$v' = v \frac{\ln\left(\frac{z'}{z_0}\right)}{\ln\left(\frac{z}{z_0}\right)} \quad (7)$$

En donde v , v' son respectivamente las velocidades a las alturas z y z' en m/s, y z_0 es la rugosidad del terreno en metros.

2.5 Parámetros y algoritmo del modelo

Para el caso de las entradas en el modelo basado con datos experimentales se utilizaron las entradas Hora, Minuto, Segundo, Temperatura, Velocidad de viento, para un total de 5 entradas mientras que la salida corresponde a la variable Densidad de potencia, es decir, 1 salida. Con respecto a la partición de los datos se utilizó un 88% para entrenamiento y 12% para validación en las zonas de Ceiba y Torcoroma, mientras que para la zona de Belén se utilizó un 84% para entrenamiento y 16% para validación. La diferencia entre las proporciones de partición de los datos es debido a que no se contaban con la misma cantidad de datos para todas las zonas y se seleccionó de tal forma que se realizara predicciones de 10 minutos para cada una de ellas. Con respecto al algoritmo de entrenamiento se utilizó el método de retropropagación del inglés Backpropagation y la técnica de validación cruzada del inglés cross-validation para evaluar los resultados del modelo, garantizando la independencia entre datos de entrenamiento y validación. Para el entrenamiento existen otros parámetros como el número de épocas, capas y neuronas. Estos últimos se han variado de tal forma de seleccionar el que resultara en un error de predicción más pequeño junto con un tiempo de ejecución aceptable, el cual se alcanzó para 300 épocas, 1 capa oculta, 64 neuronas tanto para la capa oculta como para la capa de entrada, y 1 neurona para la capa de salida. Por último, la métrica de éxito seleccionada para evaluar el modelo fue el Error Absoluto Promedio del inglés Mean Absolute Error (MAE) ya que es la métrica más utilizada y recomendada para los modelos de regresión.

3. Resultados y discusión

3.1 Perfil vertical del viento

Una vez cargadas las bases de datos a Python y realizadas las etapas de preprocesamiento y exploración de los datos donde se verificó que no hubiese valores erróneos o faltantes, se procedió a calcular el perfil vertical del viento correspondiente a la ecuación (7). En la Figura 2 se puede apreciar que la zona de Belén registró mayor perfil vertical del viento que el resto. Al tener un mayor crecimiento de la velocidad conforme aumenta la altura, también debería tener una mayor predicción de densidad de potencia, al ser ésta última proporcional a la velocidad. Esto se verificará a continuación.

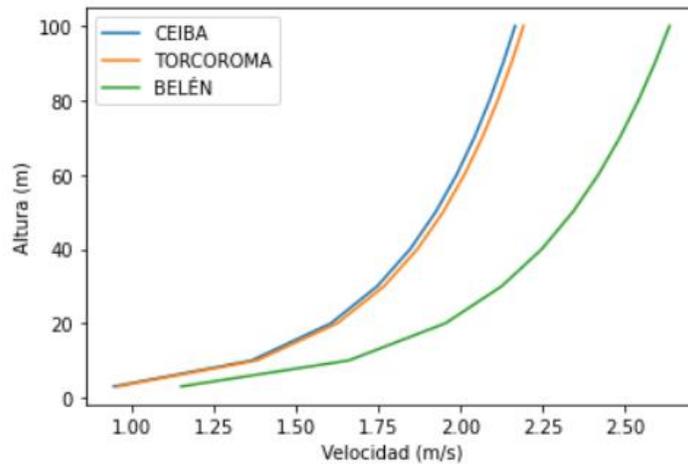


Figura 2. Perfil vertical del viento de las zonas de Ceiba, Torcoroma y Belén. Elaboración y fuente propias

3.2 Evaluación del modelo

En la Tabla 2 se aprecia la métrica MAE para 300 épocas una vez se ha entrenado el modelo. En esta comparación se observa que la zona de Ceiba registró un MAE más bajo, lo que quiere decir que sus predicciones son las más precisas, sin embargo, están todas en un rango relativamente cercano.

Tabla 1. MAE en las distintas zonas

SECTORES	MAE
CEIBA	0.0248
TORCOROMA	0.0331

BELEN	0.0606
-------	--------

Elaboración y fuente propias

3.3 Predicciones

En las Figuras 3, 4 y 5 se presentan las gráficas de los valores reales y valores predichos para las zonas de Ceiba, Torcoroma y Belén respectivamente.

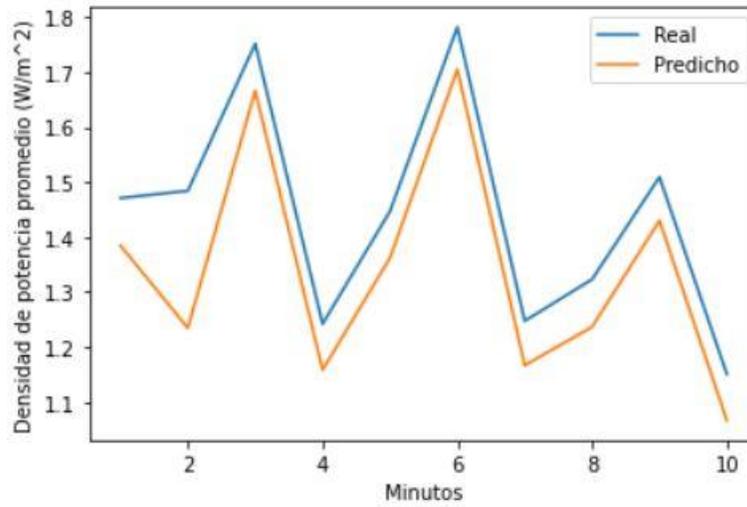


Figura 3. Densidad de potencia promedio real y predicha para la zona de Ceiba. Elaboración y fuente propias

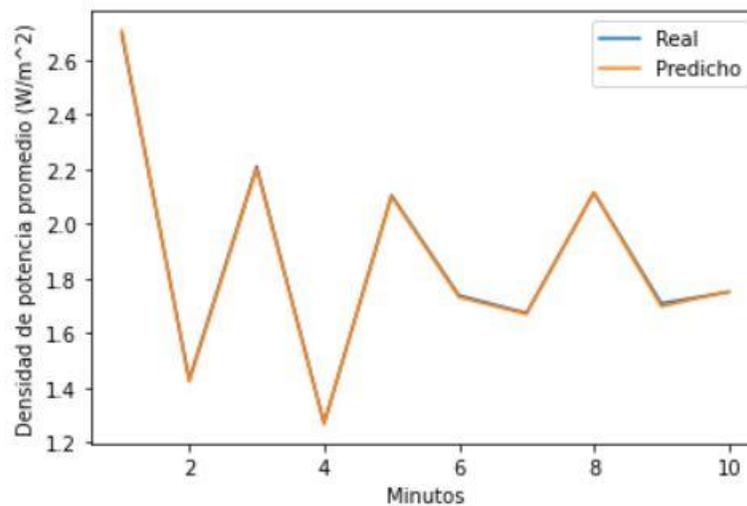


Figura 4. Densidad de potencia promedio real y predicha para la zona de Torcoroma. Elaboración y fuente propias

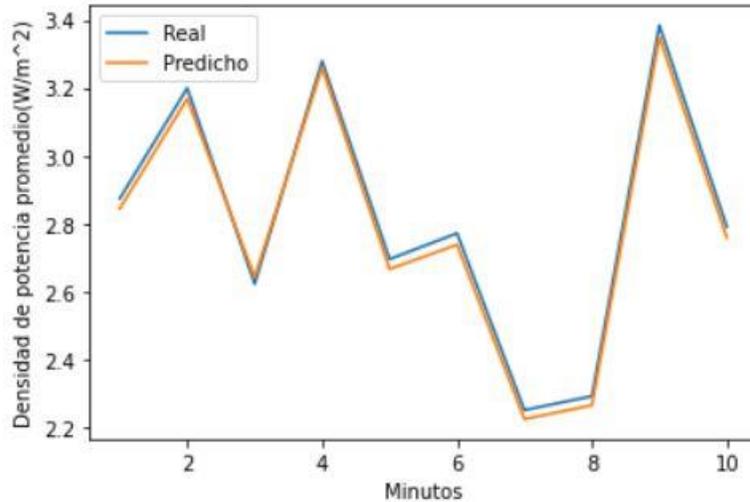


Figura 6. Densidad de potencia promedio real y predicha para la zona de Belén. Elaboración y fuente propias

Por último, se muestra en la Figura 6 la densidad de potencia promedio predicha para un pronóstico de 10 minutos en todas las zonas. Belén fue la zona que registró mayor Recurso Eólico disponible como se esperaba, ya que en el perfil de viento también fue la que tenía un mejor crecimiento de velocidad con respecto a la altura.

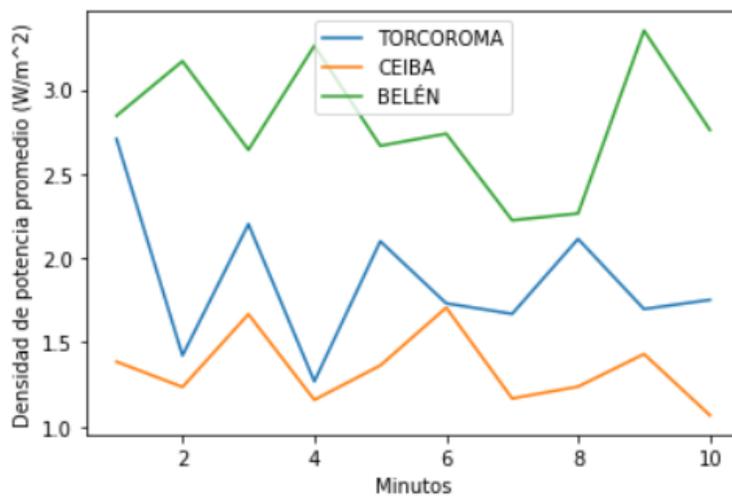


Figura 6. Densidad de potencia promedio predicha para todas las zonas. Elaboración y fuente propias

3.4 Recurso Eólico disponible promedio y total

Con los valores predichos del modelo se construyó una tabla para cada zona con la densidad de

potencia eólica disponible y sus frecuencias de aparición. Se pudo comprobar nuevamente que la zona de Belén registró mayor densidad de potencia eólica promedio y total en todo el rango de predicción.

Tabla 2. Resultados del Recurso Eólico disponible para Ceiba.

Datos			Cálculos		
			Frecuencia Relativa	Densidad de potencia eólica disponible (W/m^2)	
Límite Inferior (W/m^2)	Límite Superior (W/m^2)	Frecuencia n_i (datos)	$f_i = n_i/N$	Centro de la Clase Pd (W/m^2)	$Pd * f_i$
0	3	1074	0,87960688	1,5	1,319410319
3	6	85	0,06961507	4,5	0,313267813
6	9	39	0,031941032	7,5	0,23955774
9	12	9	0,007371007	10,5	0,077395577
12	15	6	0,004914005	13,5	0,066339066
15	18	4	0,003276003	16,5	0,054054054
18	21	1	0,000819001	19,5	0,015970516
21	24	1	0,000819001	22,5	0,018427518
24	27	0	0	25,5	0
27	30	0	0	28,5	0
30	33	0	0	31,5	0
33	36	1	0,000819001	34,5	0,028255528

36	39	0	0	37,5	0
39	42	1	0,000819001	40,5	0,033169533
Total		N=1221	1		2,165847
Resultados					
Densidad media de potencia eólica disponible					2,165847 W/m^2
Densidad de potencia eólica disponible para 10 minutos					2644,5 W/m^2

Nota: Esta tabla es una adaptación de Villarubia, M. 2012, p. 100.

Tabla 3. Resultados del Recurso Eólico disponible para Torcoroma

Datos			Cálculos		
			Frecuencia Relativa	Densidad de potencia eólica disponible (W/m^2)	
Límite Inferior (W/m^2)	Límite Superior (W/m^2)	Frecuencia ni (datos)	$f_i = ni/N$	Centro Clase Pd (W/m^2)	$Pd * f_i$
0	2,5	931	0,775187344	1,25	0,96898418
2,5	5	131	0,10907577	3,75	0,409034138
5	7,5	75	0,06244796	6,25	0,39029975
7,5	10	31	0,025811823	8,75	0,225853455
10	12,5	8	0,006661116	11,25	0,074937552
12,5	15	8	0,006661116	13,75	0,091590341

15	17,5	4	0,003330558	16,25	0,054121565
17,5	20	4	0,003330558	18,75	0,06244796
20	22,5	2	0,001665279	21,25	0,035387177
22,5	25	0	0	23,75	0
25	27,5	3	0,002497918	26,25	0,065570358
27,5	30	0	0	28,75	0
30	32,5	1	0,000832639	31,25	0,026019983
32,5	35	0	0	33,75	0
35	37,5	1	0,000832639	36,25	0,030183181
37,5	40	1	0,000832639	38,75	0,032264779
40	42,5	0	0	41,25	0
42,5	45	0	0	43,75	0
45	47,5	0	0	46,25	0
47,5	50	0	0	48,75	0
50	52,5	0	0	51,25	0
52,5	55	0	0	53,75	0
55	57,5	0	0	56,25	0
57,5	60	0	0	58,75	0
60	62,5	0	0	61,25	0
62,5	65	0	0	63,75	0
65	67,5	0	0	66,25	0
67,5	70	0	0	68,75	0
70	72,5	0	0	71,25	0

72,5	75	0	0	73,75	0
75	77,5	0	0	76,25	0
77,5	80	0	0	78,75	0
80	82,5	1	0,000832639	81,25	0,067651957
Total		1201	1		2.534346
Resultados					
Densidad media de potencia eólica disponible					2,534346 W/m^2
Densidad de potencia disponible para 10 minutos					3043,75 W/m^2

Nota: Esta tabla es una adaptación de Villarubia, M. 2012, p. 100.

Tabla 4. Resultados del Recurso Eólico disponible para Belén

Datos			Cálculos		
			Frecuencia Relativa	Densidad de potencia eólica disponible (W/m^2)	
Límite Inferior (W/m^2)	Límite Superior (W/m^2)	Frecuencia ni (datos)	$f_i = ni/N$	Centro Clase Pd (W/m^2)	$Pd * f_i$
0	2	823	0,649565904	1	0,649565904
2	4	231	0,182320442	3	0,546961326
4	6	35	0,027624309	5	0,138121547

6	8	77	0,060773481	7	0,425414365
8	10	23	0,018153118	9	0,163378058
10	12	16	0,012628256	11	0,138910813
12	14	12	0,009471192	13	0,123125493
14	16	12	0,009471192	15	0,142067877
16	18	12	0,009471192	17	0,16101026
18	20	6	0,004735596	19	0,089976322
20	22	2	0,001578532	21	0,033149171
22	24	6	0,004735596	23	0,108918706
24	26	0	0	25	0
26	28	1	0,000789266	27	0,021310182
28	30	1	0,000789266	29	0,022888713
30	32	1	0,000789266	31	0,024467245
32	34	4	0,003157064	33	0,10418311
34	36	1	0,000789266	35	0,027624309
36	38	0	0	37	0
38	40	1	0,000789266	39	0,030781373
40	42	0	0	41	0
42	44	0	0	43	0
44	46	0	0	45	0
46	48	0	0	47	0
48	50	1	0,000789266	49	0,038674033
50	52	2	0,001578532	51	0,08050513

Total		1267	1		3.071034
Resultados					
Densidad media de potencia eólica disponible					3,071034 W/m^2
Densidad de potencia disponible para 10 minutos					3891 W/m^2

Nota: Esta tabla es una adaptación de Villarubia, M. 2012, p. 100.

4. Conclusiones y recomendaciones

Con el modelo creado para analizar y predecir el Recurso Eólico disponible en las 3 zonas de Cúcuta seleccionadas se pudo identificar que Belén tuvo una mayor densidad de potencia media y total para un pronóstico de 10 minutos, lo que se corresponde también con su perfil vertical de viento, el cual tuvo un crecimiento más rápido que el resto de las zonas. Aunque en este caso se utilizó el modelo para la predicción a corto plazo, muy bien podría funcionar para rangos significativamente más amplios de datos y de tiempo, consiguiendo pronósticos más reales en el rango de los días, meses o años. De hecho, el enfoque del Aprendizaje Automático y las Redes Neuronales Artificiales suelen aumentar su precisión cuando se tienen más datos. Se recomienda que, si el tiempo seleccionado para tomar las mediciones se encuentra en el rango de los días, éstas se realicen a distintas horas del día, al igual que si se toma en el rango de los meses las mediciones se tomen distintos días de la semana, o si es en el rango de un año tomar las medidas a lo largo de cada mes; esto con el fin de cubrir las variaciones meteorológicas presentes que generan el Recurso Eólico y hacer las predicciones más reales y precisas. Por último, dentro de las recomendaciones técnicas se aconseja tomar las mediciones de todas las zonas seleccionadas a una misma altura del suelo, con el fin de aprovechar todos los datos obtenidos en las mismas condiciones.

5. Trabajo futuro

Como trabajo futuro se sugiere aplicar las predicciones a mediano y largo plazo en otras zonas de estudios que tengan condiciones meteorológicas teóricamente favorables (como zonas altas y despejadas) con el fin de obtener un acercamiento más real de las posibilidades del recurso en un

radar más amplio.

Por otra parte, también queda para trabajo futuro identificar cuanto potencial eólico es realmente aprovechable; es decir cuánta energía se puede extraer del recurso disponible. Este estudio tendría un significado relevante para analizar la viabilidad del uso de aerogeneradores de distintas escalas y potencias según pueda aplicarse.

6. Referencias

- Giraldo, M., Vacca Ramírez, R., & Urrego Quintanilla, A. (2018). Las Energías Alternativas ¿Una Oportunidad Para Colombia? *Punto de Vista*, 9(13). <https://doi.org/10.15765/pdv.v9i13.1117>
- Google Research. (s.f.). Colaboratoy. Preguntas Frecuentes. <https://research.google.com/colaboratory/faq.html>
- Guerrero, B., Vélez, F. J., & Morales, D. (2020). Energía eólica y territorio: sistemas de información geográfica y métodos de decisión multicriterio en La Guajira (Colombia). *Ambiente y Desarrollo*, 23(44). <https://doi.org/10.11144/javeriana.ayd23-44.eets>
- Norman, Alan. (2019). *Aprendizaje Automático En Acción: Un Libro Para El Lego, Guía Paso A Paso Para Los Novatos*. Italia: Tekttime.
- Palmer Pol, A., & Montaña Moreno, J. J. (s.f.). *¿Qué son las redes neuronales artificiales? Aplicaciones realizadas en el ámbito de las adicciones*. Retrieved November 11, 2021, from <http://www.disi.unal.edu.co/~lctorress/RedNeu/LiRna001.pdf>
- Rojas, E. M. (2020). *Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo*.
- Villarrubia, Miguel. (2012). *Ingeniería de la Energía Eólica*. Marcombo, S.A.

**Anexo 2. Certificado de participación en VI Congreso de
Investigación, Desarrollo e Innovación de la Universidad
Internacional de Ciencia y Tecnología IDI-UNICyT 2021**



Otorga este certificado a:

Gian Franco Di Candia Jaimes (Universidad Francisco de Paula Santander)
Brayan Orlando Calderón Angarita (Universidad Francisco de Paula Santander)
Sergio Castro (Universidad Francisco de Paula Santander)
Byron Medina (Universidad Francisco de Paula Santander)
Dinael Guevara (Universidad Francisco de Paula Santander)

por la presentación en modalidad Oral del trabajo:

**ANÁLISIS DEL RECURSO EÓLICO DISPONIBLE EN LA CIUDAD DE CÚCUTA USANDO
APRENDIZAJE AUTOMÁTICO**

en el

**VI Congreso de Investigación, Desarrollo e Innovación de la Universidad Internacional de Ciencia y Tecnología
IDI-UNICYT 2021**

del 3 al 4 de Diciembre de 2021, congreso dado en modalidad virtual síncrona.

William Núñez Alarcón
RECTOR

Universidad Internacional de Ciencia y Tecnología

Aura L. López de Ramos
Coordinadora del IDI-UNICYT

Universidad Internacional de Ciencia y Tecnología



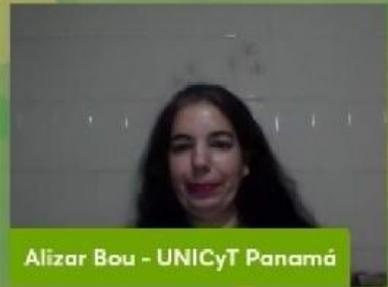
Escanee este código QR para
verificar este certificado.

Número Identificación: 1127058282
Código de verificación: 9246444D-
AEC3BD3C50EAAB408658C4752676CC66
www.esimposio.com/verificacion

Anexo 3. Exposición del proyecto en el VI Congreso de Investigación, Desarrollo e Innovación de la Universidad Internacional de Ciencia y Tecnología IDI-UNICyT 2021

Análisis del Recurso Eólico disponible en la ciudad de Cúcuta usando Aprendizaje Automático

**Analysis of available Wind Resource in city of
Cúcuta using Machine Learning**



Alizar Bou - UNICyT Panamá



Gian Di Candia - UFPS

Anexo 4. Artículo: Predicción del recurso eólico disponible en la ciudad de Cúcuta usando aprendizaje automático

Gian F. Candia-Jaimes, Brayan O. Calderon-Angarita, Sergio A. Castro-Casadiego, Byron Medina-Delgado and Dinael Guevara- Ibarra “Predicción del recurso eólico disponible en la ciudad de Cúcuta usando aprendizaje automático” sometido a revisión para aceptación en la revista EIA de la universidad EIA marzo 2022.

Predicción del recurso eólico disponible en la ciudad de Cúcuta usando aprendizaje automático

Gian F. Di Candia-Jaimes¹
Brayan O. Calderón-Angarita¹
Sergio A. Castro-Casadiegos¹
Byron Medina-Delgado¹
Dinael Guevara Ibarra¹

1. Universidad Francisco de Paula Santander

Resumen

El presente trabajo tuvo como propósito la predicción del Recurso Eólico disponible promedio y total en la ciudad de Cúcuta para un rango de un año a partir de los datos de velocidad de viento y temperatura. Los datos de velocidad de viento y temperatura fueron tomados de la página oficial del Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM, los cuales son medidos desde la torre meteorológica del Aeropuerto Camilo Daza en la ciudad de Cúcuta. Para la velocidad del viento se utilizó un rango de 4 años y una frecuencia de 2 minutos, para un total de 891478 datos. En el caso de la temperatura se tomó los valores promedios anuales para el mismo rango. Para la construcción del modelo se utilizó el enfoque de Aprendizaje Automático en conjunto con la técnica de Redes Neuronales Artificiales, el lenguaje Python y el entorno de programación de Google Colaboratory. La Red Neuronal Artificial fue entrenada utilizando el algoritmo Retropropagación con validación cruzada para evaluar los resultados del modelo, garantizando la independencia entre los datos de entrenamiento y validación para 300 épocas. Las entradas del modelo fueron las variables Año, Mes, Día, Hora, Minuto, Temperatura y Velocidad de viento, mientras que la salida fue la variable Densidad de Potencia Eólica. La métrica de éxito utilizada para evaluar el modelo fue el Error Absoluto Medio del inglés MAE, el cual tuvo un valor de 1,6, y considerando que este valor representa el 1,7% del valor promedio de las predicciones se considera que el modelo es lo suficientemente preciso para estimar el Recurso Eólico disponible

Palabras Claves: *Entrenamiento, Épocas, Error Absoluto Medio, Métrica de Éxito, Recurso Eólico, Redes Neuronales Artificiales, Retropropagación, Torre Meteorológica, Validación, Validación Cruzada*

Prediction of available Wind Resource in the city of Cúcuta using machine learning

Abstract

The present work had as purpose the prediction of the average and total available Wind Resource in the city of Cúcuta for a range of one year from the data of wind speed and temperature. The wind speed and temperature data were taken from the official website of the Institute of Hydrology, Meteorology and Environmental Studies IDEAM, which are measured from the meteorological tower of the Camilo Daza Airport in the city of Cucuta. For wind speed, a range of 4 years and a frequency of 2 minutes were used, for a total of 891.478 data. In the case of temperature, the annual average values were taken for the same range. For the construction of the model, the Machine Learning approach was used in conjunction with the Artificial Neural Networks technique, the Python language and the Google Collaborative programming environment. The Artificial Neural Network was trained using the Backpropagation algorithm with Cross-Validation to evaluate the results of the model, guaranteeing the independence between the training and validation data for 300 epochs. The inputs of the model were the variables Year, Month, Day, Hour, Minute, Temperature and Wind speed, while the output was de variable Power Density. The success metric used to evaluate the model was the Mean Absolute Error, which had a value of 1,6, and considering that this value represents 1,7% of the average value of the predictions, the model is considered to be sufficiently accurate to estimate the available Wind Resource.

Key Words: *Training, Epochs, Mean Absolute error, Success Metrics, Wind Resource, Artificial Neural Networks, Backpropagation, Meteorological Tower, Validation, Cross-Validation.*

Introducción

Antes de la llegada de la revolución industrial el 100% de la energía utilizada por el ser humano provenía que recursos que se renuevan continuamente como la leña, los alimentos, el viento y el agua (Badii, Guillen y Abreu, 2016). Sin embargo, desde los inicios de la revolución industrial la generación de energía eléctrica empezó a generarse por medio de la explotación de combustibles fósiles como el carbón, el petróleo y el gas natural. Hoy en día se conoce que esta forma de conversión de energía puede afectar en distintas medidas a la salud de las personas (Tortajada, et al., 2001). Además, este proceso también agota las reservas de combustibles, contribuye al efecto invernadero, la lluvia ácida, la deforestación e incluso compromete la eficiencia energética (Schallenberg, et al., 2008). Por estas razones muchos países han pasado a formar parte del desarrollo de energías renovables, tanto así que para el año 2016 la proporción de recursos renovables sobre el uso total de energías totales alcanzo el 10,2%, respecto al 8,6% en 2010 (Guerrero, Vélez y Morales, 2020; Gavilanez, et al., 2020).

Las energías renovables son llamadas de esta forma porque a diferencia de los combustibles fósiles no tienen que ser explotadas ni sometidas a un proceso químico para generar energía, sino que provienen de fuentes naturales e inagotables como el viento, el agua, la luz solar o la masa arbórea (Giraldo, Vacca y Urrego, 2018; Roldán, 2013). La energía derivada del movimiento de las masas de aire o del viento se conoce como energía eólica, la cual se produce por la energía solar que se mantiene en la atmósfera por medio del calor y que se distribuye en forma no homogénea. Este tipo de energía se considera autosustentable, limpia y no contaminante (González, 2009).

Para realizar la conversión de la energía solar a energía eléctrica es necesario el uso de aerogeneradores, los cuales reciben la potencia eólica disponible en una región específica para luego ser transformada a mecánica o eléctrica. La potencia eólica extraída es proporcional al cubo de la velocidad del viento, lo que hace que este parámetro sea el más determinante en el momento de estimar el Recurso Eólico disponible. Sin embargo, otros parámetros meteorológicos pueden en alguna medida ser influyentes y algunos de estos son: temperatura, dirección del viento, humedad, presión (Santamarta, 2004). En este caso se han utilizado las variables velocidad de viento y temperatura, al ser las más determinantes para estimar el Recurso Eólico disponible.

Diferentes estudios a nivel internacional han trabajado en el análisis y predicción de los datos por medio diversos lenguajes y técnicas de Aprendizaje Automático (Sandoval, 2018; Ibarguengoytia, et al., 2018; García, et al., 2019; Giraldo, Vacca y Urrego 2018; Bolaños, et al., 2020). Sin embargo, a nivel local, pocos han sido los que han basado su desarrollo en el lenguaje de programación Python, siendo este último tan efectivo para el procesamiento de datos y resultados, además de su gran variedad de bibliotecas y su fácil uso.

Teniendo en cuenta los serios inconvenientes provocados por el uso de energías convencionales, las ventajas que ofrece el Recurso Eólico y el lenguaje de programación Python, se ha propuesto realizar un modelo que realice predicciones a largo plazo del Recurso Eólico disponible en la ciudad de Cúcuta mediante el uso de Redes Neuronales Artificiales, con el fin de estudiar a futuro la viabilidad del aprovechamiento de este recurso en la ciudad.

Materiales y Métodos

Técnicas de Aprendizaje Automático

Dentro de los enfoques de la Inteligencia Artificial se encuentra el Aprendizaje Automático, definido el campo de estudio que da a la computadora la capacidad de aprender por sí misma, sin haber sido programada explícitamente (Pineda, 2022). Esto quiere decir que la máquina ya no necesita un conjunto de reglas creada por humanos para que la computadora las siga, sino que por medio de un algoritmo de aprendizaje el computador es capaz de ser entrenado para tomar decisiones por sí mismo, reconocer patrones y realizar predicciones fiables. Sin embargo, estas predicciones se realizan a expensas de no estar al tanto de qué forma se realizan las mismas (método de caja negra), ya que el mismo modelo reasignará los pesos aleatorios de las variables sin mostrarlas (Norman, 2022; Aguilar, Torres, Martín, 2019). Es importante tener en cuenta que las primeras predicciones que suelen arrojar estos

modelos son imprecisas, pero a medida que se va aumentando el número de iteraciones la computadora actualizará y ajustará los pesos de las variables, mejorando de esta forma los pronósticos.

Para llevar a cabo el enfoque del Aprendizaje Automático se consultaron diversas técnicas de Inteligencia Artificial como Redes Neuronales Artificiales, Redes Bayesianas y Máquinas de Vectores de Soporte. Para seleccionar la técnica más apropiada a las necesidades del proyecto se utilizó una matriz de Pugh donde se compararon las características más importantes a tener en cuenta para el proyecto. En la Tabla 1 se observa los parámetros seleccionados para realizar la comparación.

Tabla 1. Parámetros de comparación técnicas de Inteligencia Artificial. Fuente: Elaboración propia.

PARÁMETRO	DENOTACIÓN
Aplicaciones de predicción de tiempo	A
Habilidad con funciones no lineales	B
Cotas de error	C
Capacidad de entrenamiento	D

Teniendo ya seleccionados los parámetros se procedió a darle un peso específico a cada uno de estos, ver Tabla 2. Donde 10 es mucho más eficiente al comparado, 5 es más eficiente al comparado, 1 es igual al comparado, 0,2 es menos eficiente al comparado y 0,1 es mucho menos eficiente al comparado (Vidal, 2012).

Tabla 2. Peso de cada parámetro. Fuente: Elaboración propia.

	A	B	C	D	Suma	Factor
A	X	1,00	5,00	10,00	16,00	0,42328
B	1,00	X	5,00	10,00	16,00	0,42328
C	0,20	0,20	X	5,00	5,40	0,14286
D	0,10	0,10	0,20	X	0,40	0,01058
TOTAL					37,80	1,00

Una vez ya obtenidos los pesos de cada parámetro se procede a comparar las técnicas de Inteligencia Artificial respecto a cada parámetro, como se muestra en la Tabla 3, Tabla 4, Tabla 5 y Tabla 6.

Tabla 3. Comparación de las técnicas de Inteligencia Artificial respecto al parámetro aplicaciones de predicción de tiempo (A). Fuente: Elaboración propia.

	Red Neuronal Artificial	Red Bayesiana	Máquina de Vector de Soporte	Suma	Factor
Red Neuronal Artificial	X	5,00	10,00	15,00	0,73171
Red Bayesiana	0,20	X	5,00	5,20	0,25366
Máquina de Vector de Soporte	0,10	0,20	X	0,30	0,01463
TOTAL				20,50	1,00

Tabla 4. Comparación de las técnicas de Inteligencia Artificial respecto al parámetro habilidades con funciones no lineales (B). Fuente: Elaboración propia.

	Red Neuronal Artificial	Red Bayesiana	Máquina de Vector de Soporte	Suma	Factor
Red Neuronal Artificial	X	5,00	5,00	10,00	0,80645
Red Bayesiana	0,20	X	1,00	1,20	0,09677
Máquina de Vector de Soporte	0,20	1,00	X	1,20	0,09677
TOTAL				12,40	1,00

Tabla 5. Comparación de las técnicas de Inteligencia Artificial respecto al parámetro cotas de error (C). Fuente: Elaboración propia.

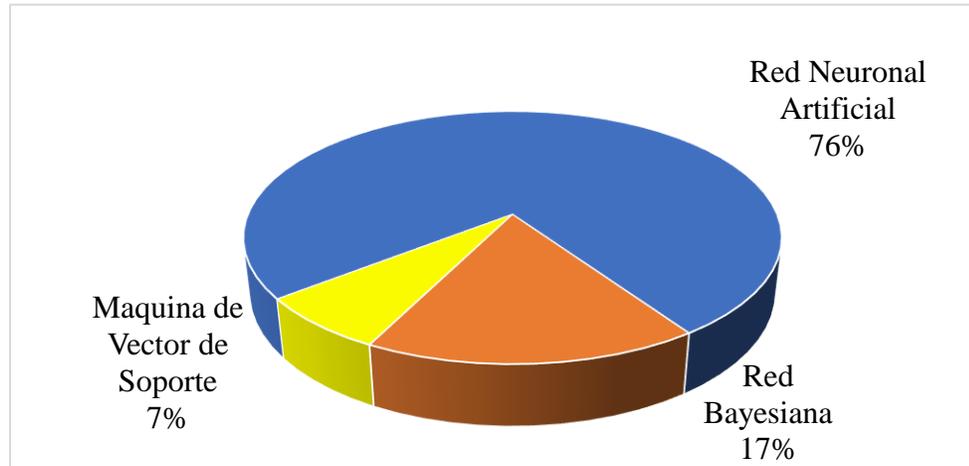
	Red Neuronal Artificial	Red Bayesiana	Máquina de Vector de Soporte	Suma	Factor
Red Neuronal Artificial	X	5,00	5,00	10,00	0,80645
Red Bayesiana	0,20	X	1,00	1,20	0,09677
Máquina de Vector de Soporte	0,20	1,00	X	1,20	0,09677
TOTAL				12,40	1,00

Tabla 6. Comparación de las técnicas de Inteligencia Artificial respecto al parámetro capacidad de entrenamiento (D). Fuente: elaboración propia.

	Red Neuronal Artificial	Red Bayesiana	Máquina de Vector de Soporte	Suma	Factor
Red Neuronal Artificial	X	0,10	0,10	0,20	0,00901
Red Bayesiana	10,00	X	1,00	11,00	0,49550
Máquina de Vector de Soporte	10,00	1,00	X	11,00	0,49550
TOTAL				22,20	1,00

Una vez realizada cada una de las comparaciones se realizó una imagen donde se logra ver los resultados obtenidos de la realización de la matriz de Pugh, en donde se multiplico el factor obtenido cada comparación con el peso de cada parámetro, ver Figura 1.

Figura 1. Selección de la técnica de Inteligencia Artificial. Fuente: Elaboración propia



Como se observa en la Figura 1 la técnica Red Neuronal Artificial fue la que tuvo un mayor porcentaje 76% es por ello que se selecciona para el presente proyecto.

Lenguaje de programación

En vista de que hoy en día son muchos los lenguajes de programación empleados en el uso de la Ciencia de los Datos y el Aprendizaje Automático, se reunió la información más relevante de cada uno de ellos (Rojas, 2020) y se seleccionó mediante una matriz de Pugh el que más se adaptaba a las necesidades del proyecto.

En la Tabla 7 se observa los parámetros elegidos para realizar la comparación con los lenguajes Python, R, Matlab y Julia. Como en la sección anterior se procede a darle un peso específico a cada parámetro Tabla 8. Seguidamente se realiza la comparación de cada lenguaje de programación respecto a cada parámetro como se observa en la Tabla 9, Tabla 10, Tabla 11, Tabla 12, Tabla 13 y Tabla 14.

Tabla 7. Parámetros de comparación lenguaje de programación. Fuente: (Rojas, 2020)

PARÁMETROS	DENOTACIÓN
Sistema operativo	A
Tipo de licencia	B
Disponibilidad de librerías	C
Especialización del software	D
Interpretación en entornos de desarrollo	E

Tabla 8. Peso de cada parámetro. Fuente: Elaboración propia.

	A	B	C	D	E	Suma	Factor
A	X	1,00	0,10	0,10	0,10	1,30	0,01662
B	1,00	X	0,10	0,10	0,10	1,30	0,01662
C	10,00	10,00	X	5,00	5,00	30,00	0,38363
D	10,00	10,00	0,20	X	0,20	20,40	0,26087
E	10,00	10,00	0,20	5,00	X	25,20	0,32225
TOTAL						78,20	1,00

Tabla 9. Comparación de los lenguajes de programación respecto al parámetro sistema operativo (A). Fuente: Elaboración propia.

	Python	R	Matlab	Julia	Suma	Factor
Python	X	1,00	1,00	1,00	3,00	0,25
R	1,00	X	1,00	1,00	3,00	0,25
Matlab	1,00	1,00	X	1,00	3,00	0,25
Julia	1,00	1,00	1,00	X	3,00	0,25
TOTAL					12,00	1,00

Tabla 10. Comparación de los lenguajes de programación respecto al parámetro tipo de licencia (B). Fuente: Elaboración propia.

	Python	R	Matlab	Julia	Suma	Factor
Python	X	5,00	10,00	1,00	16,00	0,37471
R	0,20	X	10,00	0,20	10,40	0,24356
Matlab	0,10	0,10	X	0,10	0,30	0,00703
Julia	1,00	5,00	10,00	X	16,00	0,37471
TOTAL					42,70	1,00

Tabla 11. Comparación de los lenguajes de programación respecto al parámetro disponibilidad de librerías (C). Fuente: Elaboración propia.

	Python	R	Matlab	Julia	Suma	Factor
Python	X	5,00	10,00	5,00	20,00	0,60790
R	0,20	X	5,00	1,00	6,20	0,18845
Matlab	0,10	0,20	X	0,20	0,50	0,01520
Julia	0,20	1,00	5,00	X	6,20	0,18845
TOTAL					32,90	1,00

Tabla 12. Comparación de los lenguajes de programación respecto al parámetro especialización de software (D). Fuente: Elaboración propia.

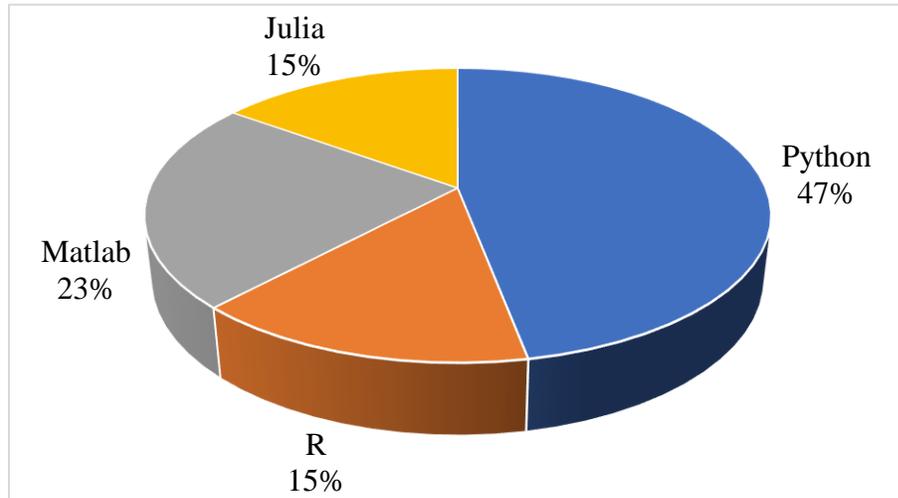
	Python	R	Matlab	Julia	Suma	Factor
Python	X	5,00	5,00	5,00	15,00	0,53571
R	0,20	X	5,00	1,00	6,20	0,22143
Matlab	0,20	0,20	X	0,20	0,60	0,02143
Julia	0,20	1,00	5,00	X	6,20	0,22143
TOTAL					28,00	1,00

Tabla 13. Comparación de los lenguajes de programación respecto al parámetro interpretación en entornos de desarrollo (E). Fuente: Elaboración propia.

	Python	R	Matlab	Julia	Suma	Factor
Python	X	5,00	0,20	5,00	10,20	0,26984
R	0,20	X	0,10	1,00	1,30	0,03439
Matlab	5,00	10,00	X	10,00	25,00	0,66138
Julia	0,20	1,00	0,10	X	1,30	0,03439
TOTAL					37,80	1,00

Una vez realizada cada comparación se reunieron los resultados de la matriz de Pugh en la Figura 2, donde se multiplicó el factor obtenido de cada comparación con el peso de cada parámetro de la Tabla 7.

Figura 2. Selección del lenguaje de programación. Fuente: Elaboración propia.



Como se observa en la Figura 2 el lenguaje de programación Python fue quien mayor puntaje 47% obtuvo respecto a los demás, por esta razón, se seleccionó al lenguaje Python para la realización del proyecto.

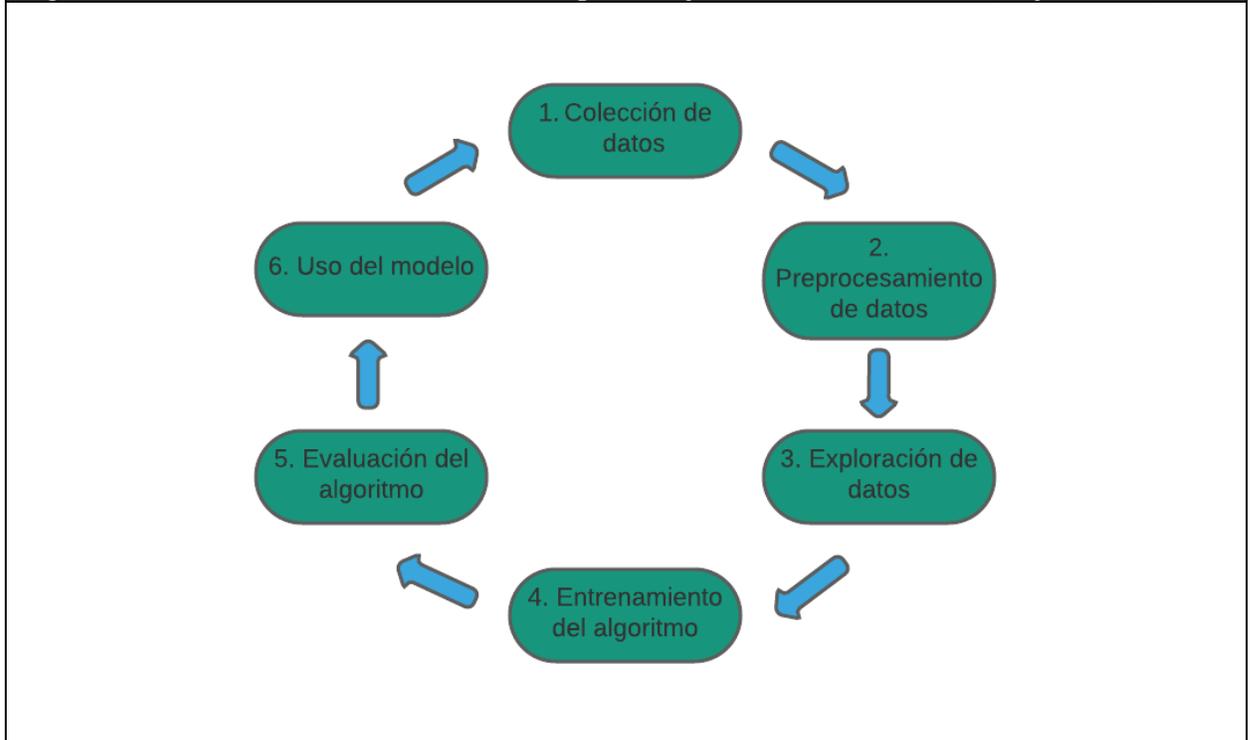
Entorno de programación

El entorno utilizado para realizar el modelo fue Google Colaboratory, el cual es un producto de Google Research y permite desarrollar código en Python en el navegador. Es útil para desarrollar aplicaciones de aprendizaje automático, análisis de datos y educación, y es que Google Colaboratory permite la entrada de texto lo que hace más dinámico su uso. Este entorno no requiere instalaciones de paquetes ni configuraciones. Además, cuenta con una GPU y otros recursos computacionales gratuitos (Google Colaboratory). La ventaja de contar con una GPU permitió acelerar significativamente los procesos de entrenamiento del algoritmo de la Red Neuronal Artificial, lo cual fue de suma importancia teniendo en cuenta que se utilizaron más de 800 mil datos.

Creación del modelo

Para el siguiente paso que es la construcción del modelo de Aprendizaje Automático es prioritario seguir una serie de pasos para garantizar el funcionamiento óptimo de las predicciones. En la Figura 3 se puede observar los pasos a seguir.

Figura 3. Proceso de creación del modelo Aprendizaje Automático. Fuente: (Rojas, 2020)



Colección de datos

Los datos de velocidad de viento y temperatura fueron tomados de la página oficial del Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM. Para la velocidad del viento se utilizó un rango de 4 años y una frecuencia de 2 minutos, para un total de 891478 datos. En el caso de la temperatura se tomó los valores promedios anuales para el mismo rango. En la Figura 4 se observa el marco de datos en Python para velocidad de viento y en la Figura 5 el marco datos correspondiente a los valores de temperatura.

Figura 4. Marco de datos correspondientes a los valores de velocidad del viento para un rango de 4 años y una frecuencia de 2 minutos. Fuente: Elaboración propia

	Fecha	VelocidadDeViento
0	01/01/2017 00:00	2.2
1	01/01/2017 00:02	2.1
2	01/01/2017 00:04	2.3
3	01/01/2017 00:06	2.1
4	01/01/2017 00:08	2
...
891473	31/12/2020 22:54	1.4
891474	31/12/2020 22:56	1.2
891475	31/12/2020 22:58	1.6
891476	31/12/2020 23:00	1.1
891477	NaN	NaN

891478 rows x 2 columns

Figura 5. Marco de datos correspondientes a los valores de temperatura promedio anual para un rango de 4 años. Fuente: Elaboración propia

	Fecha	Valor
0	01/01/2017 00:00	27.354348
1	01/01/2018 00:00	26.831481
2	01/01/2019 00:00	27.679116
3	01/01/2020 00:00	27.290892

Preprocesamiento de los datos

En este paso es necesario asegurarse que todos los datos previamente descargados se encuentren en el mismo formato para que el algoritmo se entrene de manera correcta.

Fueron necesarias diferentes sentencias y condiciones ejecutadas en Python para garantizar que el formato fuera igual para todas las variables. Para este caso se tienen dos marcos de datos, velocidad del viento y temperatura. Ambos con la variable fecha. Para establecer condiciones con respecto a las variables fue necesario que todas se encontraran en el formato de tipo numérico. Para ello de la variable Fecha que es de tipo “datetime64” en el marco de datos de velocidad de viento, se han extraído las variables numéricas Año, Mes, Día, Hora, Minuto, como se muestra en la Figura 6.

Figura 6. Conversión del tipo de dato datetime a numérico para la variable Fecha en el marco de datos de velocidad. Fuente: Elaboración propia.

```
df_WW['Fecha'] = pd.to_datetime(df_WW['Fecha'], errors='coerce', format='%d/%m/%Y %H:%M')

df_WW['Año'] = df_WW['Fecha'].dt.year
df_WW['Mes'] = df_WW['Fecha'].dt.month
df_WW['Día'] = df_WW['Fecha'].dt.day
df_WW['Hora'] = df_WW['Fecha'].dt.hour
df_WW['Minuto'] = df_WW['Fecha'].dt.minute
```

Así como la conversión del tipo de dato de la variable velocidad de viento a numérico, el cual era inicialmente de tipo objeto, ver Figura 7.

Figura 7. Conversión del tipo de dato objeto a numérico para velocidad de viento en el marco de datos de velocidad. Fuente: Elaboración propia.

```
[ ] df_WW.VelocidadDeViento=pd.to_numeric(df_WW.VelocidadDeViento, errors='coerce')
```

Para el marco de datos de temperatura se realizó el mismo procedimiento para la variable Fecha, como se observa en la Figura 8.

Figura 8. Conversión del tipo de dato datetime a numérico para la variable Fecha en el marco de datos de temperatura. Fuente: Elaboración propia.

```
df_T['Fecha'] = df_T['Fecha'].astype('datetime64[ns]')

df_T['Año'] = df_T['Fecha'].dt.year
df_T['Mes'] = df_T['Fecha'].dt.month
df_T['Día'] = df_T['Fecha'].dt.day
df_T['Hora'] = df_T['Fecha'].dt.hour
df_T['Minuto'] = df_T['Fecha'].dt.minute
```

Para la variable temperatura no fue necesario cambiar el dato ya que este estaba establecido como tipo numérico, sin embargo, se pasó de grados Celsius a grados Kelvin para facilitar cálculos posteriores, en la figura 9 se observa como quedó el marco de datos de las variables de entrada del modelo una vez convertido todas las variables a tipo numérico y unidos en un mismo marco de datos donde la temperatura está expresada en Kelvin y la velocidad de viento en (m/s).

Figura 9. Marco de datos de tipo numérico con las entradas del modelo. Fuente: Elaboración propia.

	Año	Mes	Día	Hora	Minuto	TemperaturaK	VelocidadDeViento
0	2017.0	1.0	1.0	0.0	0.0	300.504	2.2
1	2017.0	1.0	1.0	0.0	2.0	300.504	2.1
2	2017.0	1.0	1.0	0.0	4.0	300.504	2.3
3	2017.0	1.0	1.0	0.0	6.0	300.504	2.1
4	2017.0	1.0	1.0	0.0	8.0	300.504	2.0
...
891472	2020.0	12.0	31.0	22.0	52.0	300.441	2.0
891473	2020.0	12.0	31.0	22.0	54.0	300.441	1.4
891474	2020.0	12.0	31.0	22.0	56.0	300.441	1.2
891475	2020.0	12.0	31.0	22.0	58.0	300.441	1.6
891476	2020.0	12.0	31.0	23.0	0.0	300.441	1.1

890145 rows × 7 columns

Exploración de los datos

En la exploración de los datos se procedió a analizar si existen valores faltantes o atípicos que pudieran afectar el modelo. Además, también se buscó si existe algún tipo de patrón que sirva de ayuda a la construcción del mismo, en la Figura 10 se puede observar un resumen estadístico de los datos, como valores promedio, mínimos, máximos, así como los percentiles 25, 50 y 75.

Figura 10. Resumen estadístico del marco de datos. Fuente: Elaboración propia.

	Año	Mes	Día	Hora	Minuto	VelocidadDeViento
count	891430.000000	891430.000000	891430.000000	891430.000000	891430.000000	891430.000000
mean	2018.435064	6.467597	15.847223	11.259040	29.038215	4.415768
std	1.096977	3.468874	8.764593	6.813686	17.327244	23.746525
min	2017.000000	1.000000	1.000000	0.000000	0.000000	0.000000
25%	2017.000000	3.000000	8.000000	5.000000	14.000000	1.400000
50%	2018.000000	7.000000	16.000000	11.000000	30.000000	2.900000
75%	2019.000000	9.000000	23.000000	17.000000	44.000000	5.700000
max	2020.000000	12.000000	31.000000	23.000000	58.000000	999.900000

La figura anterior fue de gran utilidad para determinar que a pesar de que el 75% de los datos de velocidad de viento corresponden a un valor de 5,7 (m/s) o menor, y el valor máximo es 999,9 (m/s) lo cual es un valor atípico al ser excesivamente alto. Debido a que

los valores que superaban los 15 (m/s) eran escasos y muy distantes entre ellos se acoto el rango de datos de la variable velocidad de viento a un máximo de 15 (m/s) para evitar malos pronósticos al usar el modelo. La instrucción utilizada para acotar el rango, así como el resumen estadístico del nuevo marco de datos se presentan en la figura 11.

Figura 11. Acote del rango y resumen estadístico del nuevo marco de datos. Fuente: Elaboración propia.

```
df_VW=df_VW[df_VW['VelocidadDeViento']<=15]
```

```
df_VW.describe()
```

	Año	Mes	Día	Hora	Minuto	VelocidadDeViento
count	890145.000000	890145.000000	890145.000000	890145.000000	890145.000000	890145.000000
mean	2018.434592	6.467218	15.848755	11.258759	29.038408	3.847193
std	1.097080	3.469636	8.764616	6.817204	17.327013	2.980743
min	2017.000000	1.000000	1.000000	0.000000	0.000000	0.000000
25%	2017.000000	3.000000	8.000000	5.000000	14.000000	1.400000
50%	2018.000000	7.000000	16.000000	11.000000	30.000000	2.900000
75%	2019.000000	9.000000	23.000000	17.000000	44.000000	5.700000
max	2020.000000	12.000000	31.000000	23.000000	58.000000	15.000000

Ecuaciones

El principal factor para la generación de energía eólica es el viento, el cual es generado por el movimiento de la masa atmosférica. Sin embargo para la estimación del potencial eólico es necesario conocer no sólo la velocidad del viento sino también otros parámetros como la dirección del viento, densidad y humedad atmosférica (Chercca, 2014; Encinas, et al., 2015). A partir de los datos obtenidos de velocidad de viento y temperatura, se procedió a utilizar la ecuación 1 para calcular la densidad de potencia para cada uno de los valores obtenidos.

La forma en que se suele estimar el Recurso Eólico es mediante la potencia que se puede extraer del viento y se puede apreciar en (1).

$$E = \frac{1}{2}mv^3 \quad (1)$$

Donde E es la energía cinética derivada del viento, m es la masa de aire y v es la velocidad del viento. Esta energía cinética también se puede expresar como (2).

$$Pd = \frac{1}{2}mv^3 = \frac{1}{2}\rho Av^3 \quad (2)$$

Donde A es el área ortogonal a la dirección del flujo del aire, ρ es la densidad de la masa de aire m y Pd es el potencial eólico disponible. Si despejamos Pd/A se puede observar la

potencia disponible por unidad de área en (3).

$$\frac{Pd}{A} = \frac{1}{2} \rho v^3 \quad (3)$$

Además, la densidad de la masa de aire ρ se puede expresar en (4).

$$\rho = \frac{Patm}{Rp \cdot Tk} \quad (4)$$

Donde $Patm$ es la presión atmosférica, Rp es la constante universal sobre masa molecular del gas para el aire y Tk es la temperatura en Kelvin. Por otro lado, la presión atmosférica se calcula mediante (5).

$$Patm = Po \cdot e^{-\left(\frac{g \cdot Z}{Rp \cdot Tk}\right)} \quad (5)$$

Donde g es la gravedad en $[m/s^2]$, $Patm$ es la presión atmosférica, Po es la presión a nivel del mar es $[Pa]$ y Z es la altura sobre el nivel del mar de la zona en $[m]$. Por último la constante Rp y la variable Tk definida anteriormente como la constante universal sobre masa molecular del gas para el aire y la temperatura en Kelvin respectivamente.

Si se sustituye (5) en (4) y luego en (3) se obtiene (6).

$$\frac{Pd}{A} = \frac{Po \cdot e^{-\left(\frac{g \cdot Z}{Rp \cdot Tk}\right)} \cdot v^3}{2 \cdot Rp \cdot Tk} \left[\frac{W}{m^2} \right] \quad (6)$$

Con (6) se obtuvieron los valores de densidad de potencia para cada par temperatura Tk y velocidad v .

Por otro lado, la velocidad del viento medio o estacionario se obtiene a partir del promedio de velocidad para intervalos de tiempo de 10 minutos (Villarrubia, 2011). Su variación con respecto a la altura se corresponde con (7).

$$v' = v \frac{\ln\left(\frac{z'}{zo}\right)}{\ln\left(\frac{z}{zo}\right)} \quad (7)$$

En donde v , v' son respectivamente las velocidades a las alturas z y z' en m/s, y zo es la rugosidad del terreno en metros.

Entrenamiento del modelo

Una vez se ha limpiado el marco de datos en las etapas de preprocesamiento y exploración se puede proseguir con el entrenamiento del modelo.

Para el caso de las entradas en el modelo basado con datos experimentales se utilizaron las entradas Año, Mes, Día, Hora, Minuto, Temperatura, Velocidad de viento, para un total de 7 entradas, mientras que la salida corresponde a la variable Densidad de potencia, es decir, 1 salida. Con respecto a la partición de los datos se utilizó un 78,4% para entrenamiento y 21,6% para validación. A pesar de que en teoría 1 año sería el 25% de los 4 años totales, también es cierto que no se tiene la misma cantidad de datos por año. Con respecto al algoritmo de entrenamiento se utilizó el método de retropropagación del inglés *Backpropagation* y la técnica de validación cruzada del inglés *cross-validation* para evaluar los resultados del modelo, garantizando la independencia entre datos de entrenamiento y validación. Para el entrenamiento existen otros parámetros como el número de épocas, capas y neuronas.

Estos últimos se han variado de tal forma de seleccionar el que resultara en un error de predicción más pequeño junto con un tiempo de ejecución aceptable, el cual se alcanzó para 300 épocas, 1 capa oculta, 64 neuronas tanto para la capa oculta como para la capa de entrada, y 1 neurona para la capa de salida. Por último, la métrica de éxito seleccionada para evaluar el modelo fue el Error Absoluto Promedio del inglés Mean Absolute Error (MAE) ya que es la métrica más utilizada y recomendada para los modelos de regresión. El resumen de los parámetros del modelo se puede observar en la Tabla 14.

Tabla 14. Resumen de los parámetros utilizados para el entrenamiento del modelo. Fuente: Elaboración propia.

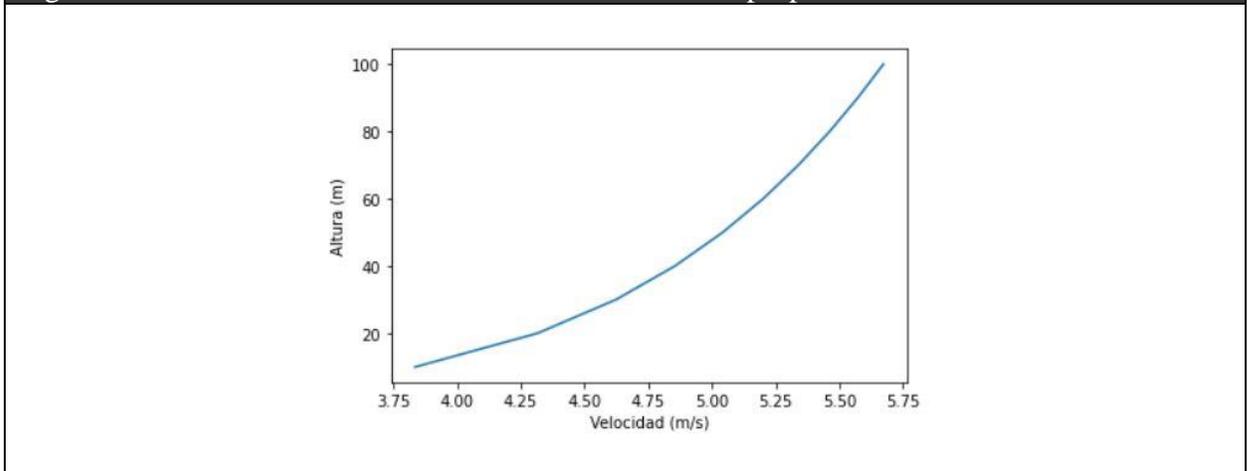
Número de entradas	Número de salidas	Partición de los datos	Algoritmo de entrenamiento	Número de capas ocultas	Neuronas en la capa oculta	Número de épocas	Métrica de éxito
7	1	78.4% entrenamiento 21,6% validación	Retropropagación	1	64	300	EAM

Resultados

Perfil vertical del viento

Con el fin de estimar la velocidad del viento medio o estacionario se tomaron los valores promedio para intervalos de 10 minutos y se utilizó la ecuación 7 para calcular la variación de la velocidad conforme varía la altura. La importancia de este análisis se basa en tener una proyección real sobre el comportamiento del viento a una altura mayor a la que tiene la torre meteorológica, con el objetivo futuro de evaluar la posibilidad del uso de aerogeneradores con alturas mayores. También se pudo observar que a medida que se aumenta la altura, también lo hace la velocidad del viento. En la Figura 12 se aprecia este comportamiento.

Figura 12. Perfil vertical del viento. Fuente: Elaboración propia.



Evaluación del modelo

Para la evaluación del modelo se tuvo en cuenta la variación del número de épocas con el fin de aumentar la precisión, es decir que el MAE disminuyera lo más posible. Se pudo observar que a medida que se aumentaba el número de épocas, el error disminuía, sin embargo, para un número de épocas mayor se requiere un tiempo de simulación mayor, lo que puede generar un gasto considerablemente mayor de tiempo, o en su defecto mejores recursos computacionales. Por esta razón se fijó el modelo para 300 épocas. En la Tabla 15 se puede apreciar los resultados del error MAE conforme se aumentaron las épocas, alcanzando un valor mínimo de 1,64 para 300 épocas. Por otra parte, en la Figura 13 se observa la relación entre el número de épocas y el MAE. Además, una comparación entre los valores reales y los valores predichos promedios mensuales para 1 año se puede apreciar en la Figura 14.

Tabla 15. Relación entre número de épocas y el MAE. Fuente: Elaboración propia.

Número de épocas	MAE
50	21,2163
100	4,0124
150	2,9001
200	2,3458
250	2,2038
300	1,6449

Figura 13. Relación entre el número de épocas y MAE. Fuente: Elaboración propia.

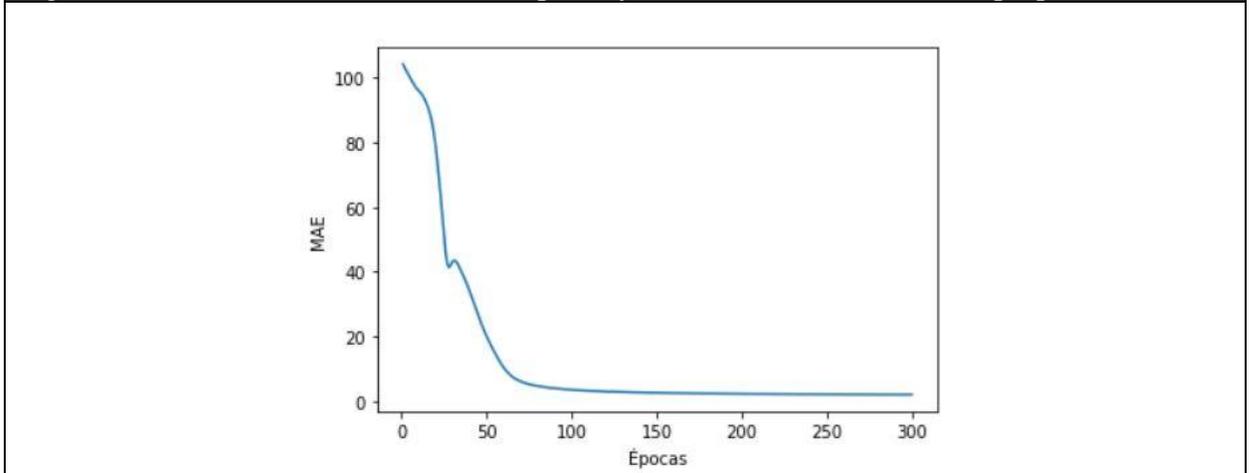
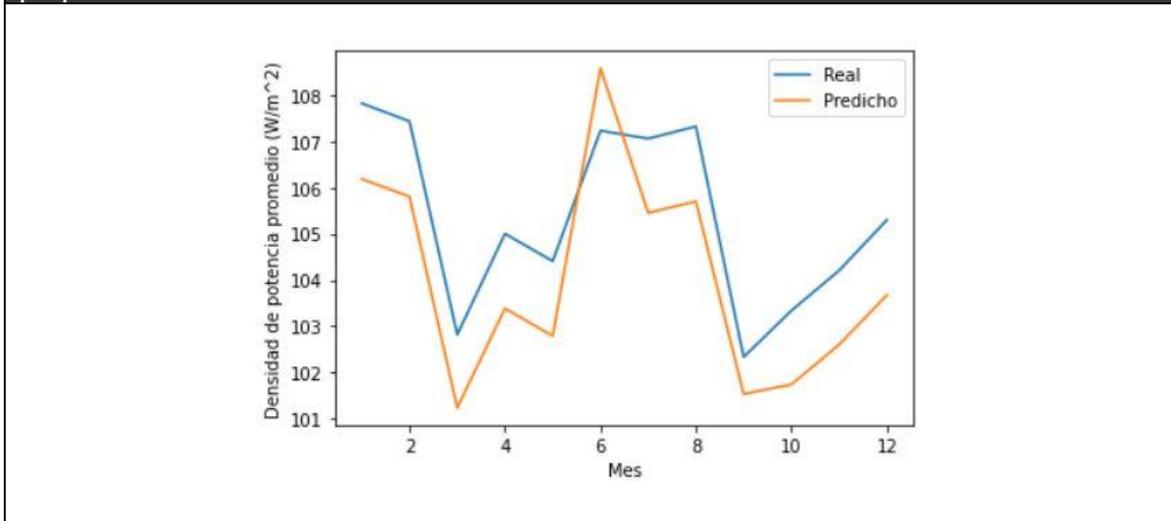


Figura 14. Valores reales vs valores predichos para el año 2020. Fuente: Elaboración propia.



Recurso Eólico disponible promedio y total

A partir de los datos de las predicciones del Recurso Eólico del último año obtenidas del modelo se construyó una tabla donde se obtuvo la frecuencia para cada rango de valor del recurso, así como el número total de datos. Para obtener el Recurso Eólico promedio anual se multiplicó cada frecuencia relativa por el centro de la clase y se sumaron todos los valores. Mientras que para obtener el Recurso Eólico disponible total del año se multiplicó el valor promedio obtenido anteriormente por el número de datos. Este procedimiento se puede apreciar en la Tabla 16.

Tabla 16. Recurso Eólico disponible promedio y total. Fuente: [8].

Datos			Cálculos		
			Frecuencia Relativa	Densidad de potencia eólica disponible (W/m^2)	
Límite Inferior (W/m^2)	Límite Superior (W/m^2)	Frecuencia ni (datos)	$f_i=ni/N$	Centro Clase P_d (W/m^2)	P_d*f_i
0	50	123594	0,642303675	25	16,05759187
50	100	18550	0,096402197	75	7,230164793
100	150	10932	0,056812335	125	7,101541915
150	200	7682	0,039922462	175	6,986430936
200	250	5096	0,026483321	225	5,958747135
250	300	4519	0,023484719	275	6,458297605
300	350	3522	0,018303425	325	5,948613212
350	400	2940	0,015278839	375	5,729564553
400	450	2677	0,013912058	425	5,91262479
450	500	1830	0,009510298	475	4,517391372
500	550	1685	0,008756749	525	4,597293463
550	600	1456	0,007566663	575	4,350831242
600	650	1261	0,006553271	625	4,095794162
650	700	997	0,005181293	675	3,497372975
700	750	872	0,004531683	725	3,285470032
750	800	675	0,003507897	775	2,718619915
800	850	687	0,003570259	825	2,9454639
850	900	464	0,002411354	875	2,109934883
900	950	431	0,002239857	925	2,071867708
950	1000	373	0,001938438	975	1,88997677
1000	1050	334	0,001735759	1025	1,77915322
1050	1100	303	0,001574656	1075	1,692755024
1100	1150	251	0,001304418	1125	1,467470105
1150	1200	125	0,00064961	1175	0,76329233
1200	1250	180	0,000935439	1225	1,145912911
1250	1300	202	0,001049771	1275	1,338457461
1300	1350	105	0,000545673	1325	0,723016479
1350	1400	121	0,000628823	1375	0,864631567
1400	1450	116	0,000602839	1425	0,859044917
1450	1500	76	0,000394963	1475	0,582570691
1500	1550	71	0,000368979	1525	0,56269261
1550	1600	73	0,000379373	1575	0,597511732

1600	1650	59	0,000306616	1625	0,498251249
1650	1700	37	0,000192285	1675	0,322076883
1700	1750	45	0,00023386	1725	0,403408116
1750	1800	33	0,000171497	1775	0,304407477
1800	1850	33	0,000171497	1825	0,312982336
1850	1900	16	8,31501E-05	1875	0,155906518
Resultados					
Densidad media de potencia eólica disponible en W/m^2					117,8371349
Densidad de potencia disponible total para 1 año en W/m^2					22.674575

Discusión

El modelo creado en el presente proyecto es capaz de realizar predicciones del Recurso Eólico a largo plazo con una precisión relativamente alta, y que pudiera incrementarse aún más con la mejora de los recursos computacionales. Aunque en este proyecto se utilizaron datos meteorológicos de una base de datos pública, muy bien pudiera funcionar con datos medidos experimentalmente en otras zonas de estudios, siempre y cuando se tenga en cuenta las consideraciones matemáticas que implica las diferencias de altura entre las mediciones realizadas y las especificadas en el estudio. Por otro lado, a pesar de que el modelo arrojó una precisión buena, es importante resaltar que con una mayor cantidad de datos se suele lograr una mejor precisión, sobre todo si se trata de una predicción a largo plazo, sin embargo, esto puede verse limitado por la cantidad que ofrece la base de datos utilizada. A pesar de ello, este proyecto destaca su validez en que no sólo la precisión es relativamente alta, sino que además las predicciones tuvieron un comportamiento fiel a las originales, en el sentido de que los picos altos y bajos coinciden casi en su totalidad.

Con respecto a la evaluación del modelo de la Red Neuronal Artificial se recomienda realizar varias veces el entrenamiento para un mismo número de épocas y luego sacar la media de los resultados obtenidos, ya que debido a que la Red trabaja con pesos aleatorios para las neuronas, va a arrojar diferentes resultados cada vez que se ejecute. La idea es verificar que estos resultados no difieran mucho unos de otros y presentar una métrica de éxito promedio que refleje la precisión del entrenamiento.

Como trabajo futuro se sugiere aplicar las predicciones a mediano y corto plazo en otras zonas de estudios que tengan condiciones meteorológicas teóricamente favorables (como zonas altas y despejadas) con el fin de obtener un acercamiento más real de las posibilidades del recurso en un radar más amplio.

Por otra parte, también queda para trabajo futuro identificar cuanto potencial eólico es realmente aprovechable; es decir cuánta energía se puede extraer del recurso eólico disponible. De esta forma, este estudio podría iniciar el proceso de análisis de viabilidad del uso de aerogeneradores de distintas escalas y potencias según pueda aplicarse.

Referencias

Aguilar, R.M., Torres, J.M., Martín, C.A. (2019). Aprendizaje Automático en la Identificación de Sistemas. Un Caso de Estudio en la Predicción de la Generación Eléctrica de un Parque Eólico. *Revista Iberoamericana de Automática e Informática Industrial*, 16(1). DOI: <https://doi.org/10.4995/riai.2018.9421>

Badii, M.H., Guillen, A., Abreu, J.L. (2016). Energías Renovables y Conservación de Energía. *Daena: Internacional Journal of Good Conscience*, 11(1)141-155. ISSN 1870-557X

Bolaños-Logroño, P. F., Yaulema-Castañeda, J. L., Gavilánez-Álvarez, O. D., & Ayavaca-Vallejo, B. L. (2020). Redes neuronales en la predicción de la velocidad del viento. *Polo del Conocimiento*, 5(5). DOI: <http://dx.doi.org/10.23857/pc.v5i5.1398>. Recuperado de: https://redib.org/Record/oai_articulo2661736-redes-neuronales-en-la-predicci%C3%B3n-de-la-velocidad-del-viento

Chercca Ramírez, J. A. (2014). Aprovechamiento del recurso eólico y solar en la generación de energía eléctrica y la reducción de emisiones de CO2 en el poblado rural La gramita de Casma. Recuperado de: <http://cybertesis.uni.edu.pe/handle/uni/1694>

Encinas, Jesus, Torrez, Rene, Ghezzi, Flavio, & Palenque, Eduardo R.. (2015). Características del viento y del potencial eólico del altiplano central. *Revista Boliviana de Física*, 27(27), 20-24. Recuperado el 09 de marzo de 2022, de: http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S1562-38232015000200004&lng=es&tlng=es

García, Uriel & Ibarquengoytia, Pablo & González, Lorena & Valadez, Jorge. (2019). Técnicas de aprendizaje automático en el diagnóstico de aerogeneradores. *Revista de Energías Renovables*. 7-14. DOI: 10.35429/JRE.2019.11.3.7.14.

Gavilanez, A., Caiza, G., Tapia, M. J., & Mora-Aguilar, J. (2020). Energías Renovables y Diseño Industrial: Influencia en Sudamérica. *INGENIO*, 3(2), 55-67. <https://doi.org/10.29166/ingenio.v3i2.2722>

Giraldo, M., Vacca Ramírez, R., & Urrego Quintanilla, A. (2018). LAS ENERGÍAS ALTERNATIVAS ¿UNA OPORTUNIDAD PARA COLOMBIA? *Punto De Vista*, 9(13). <https://doi.org/10.15765/pdv.v9i13.1117>

González Velasco, J. (2009). *Energías renovables*. España, Reverté. ISBN: 9788429179125

Google Colaboratory. Disponible en: <https://colab.research.google.com/>

Guerrero Hoyos, B. G., Vélez Macías, F. de J., & Morales Quintero, D. E. (2020). Energía eólica y territorio: sistemas de información geográfica y métodos de decisión multicriterio en La Guajira (Colombia). *Ambiente Y Desarrollo*, 23(44). <https://doi.org/10.11144/Javeriana.ayd23-44.eets>

Ibargüengoytia-González, Pablo Héctor, Reyes-Ballesteros, Alberto, Borunda-Pacheco, Mónica, & García-López, Uriel Alejandro. (2018). Predicción de potencia eólica utilizando técnicas modernas de Inteligencia Artificial. *Ingeniería, investigación y tecnología*, 19(4), e033. <https://doi.org/10.22201/fi.25940732e.2018.19n4.033>

i Tortajada, J. F., i Castell, J. G., Andreu, J. L., i Domínguez, F. G., García, J. O., Tornero, O. B., ... & Conesa, A. C. (2001). Enfermedades asociadas a la polución atmosférica por combustibles fósiles. Aspectos pediátricos. *Revista Española de Pediatría*, 57(3), 213-225.

Recuperado de: <https://pesquisa.bvsalud.org/portal/resource/pt/ibc-471>

Norman, A.T., Aprendizaje Automático En Acción. (2022). Rusia: ЛитРес.

Pineda Pertuz, C. (2022). Aprendizaje automático y profundo en python: Una mirada hacia la inteligencia artificial. Colombia: Ediciones de la U. pp 35-36. ISBN: 9785042842290, 5042842292

Rojas, E. M. Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. *Revista Ibérica de Sistemas e Tecnologias de Informação; Lousada N.º E28, 04/2020*. ISSN: 1646-9895

ROLDÁN VILORIA, J. (2013). *Energías renovables. Lo que hay que saber*. España, Ediciones Paraninfo, S.A. ISBN 13: 978-84-283-2968-2

Sandoval Serrano, L. (2018). Algoritmos de aprendizaje automático para análisis y predicción de datos. *Revista Tecnológica*. ISSN 2072-568X. <http://hdl.handle.net/10972/3626>

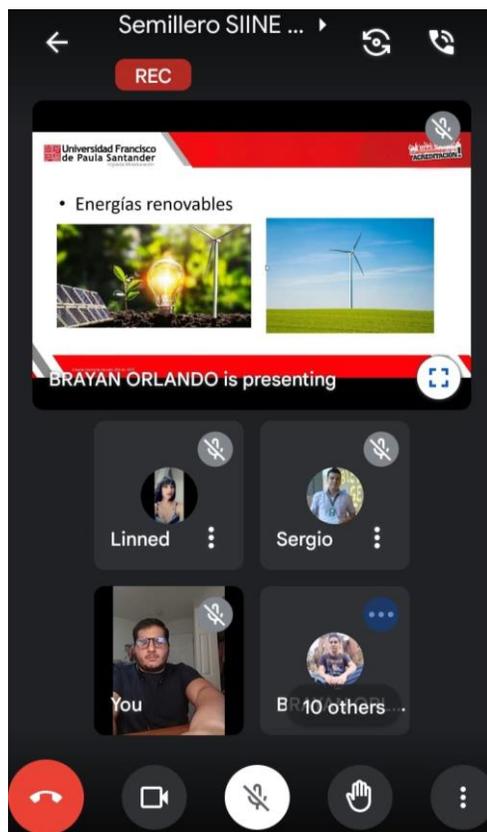
Santamarta Flórez, José. (2004). Las energías renovables son el futuro. *World Watch*. ISSN 1136-8586, N° 22, 2004, págs. 34-40

Schallenberg-Rodríguez, J.C., Piernavieja-Izquierdo, G., Hernández-Rodríguez, C. Unamunzaga-Falcón, P., García-Déniz, R., Días-Torres, M., Cabrera-Pérez, D., Martel-Rodríguez, G., Pardilla-Fariña, J., Subiela-Ortín, V. (2008). *Energías renovables y eficiencia energética*. España, Instituto Tecnológico de Canarias, S.A., pp 19-20. Recuperado de: <http://up-rid2.up.ac.pa:8080/xmlui/handle/123456789/1459>

Vidal-Holguín, C.J., Bravo-Bastidas, J.J., Cajiao-Gómez, E., Meza-Herrera, P.P., Arango-Sanclemente, S. Franco-Leyton, D., Calderón Sotero, J.H. (2012). *Guía metodológica para la priorización de proyectos: Un enfoque aplicado a la infraestructura, la logística y la conectividad*. Cali, Sello Editorial Javeriano-Pontificia Universidad Javeriana. ISBN: 9588347653, 9789588347653.

Villarrubia López, M. (2011). *Ingeniería de la Energía Eólica*. España: Marcombo, pp 97-100. ISBN: 9788426715807

Anexo 5. Divulgación en el semillero SINE mediante la plataforma Google Meet



Anexo 6. Manual de usuario de la Interfaz desarrollada para la predicción del Recurso Eólico disponible en la ciudad de Cúcuta

Interfaz para la predicción del Recurso Eólico disponible en la ciudad de Cúcuta

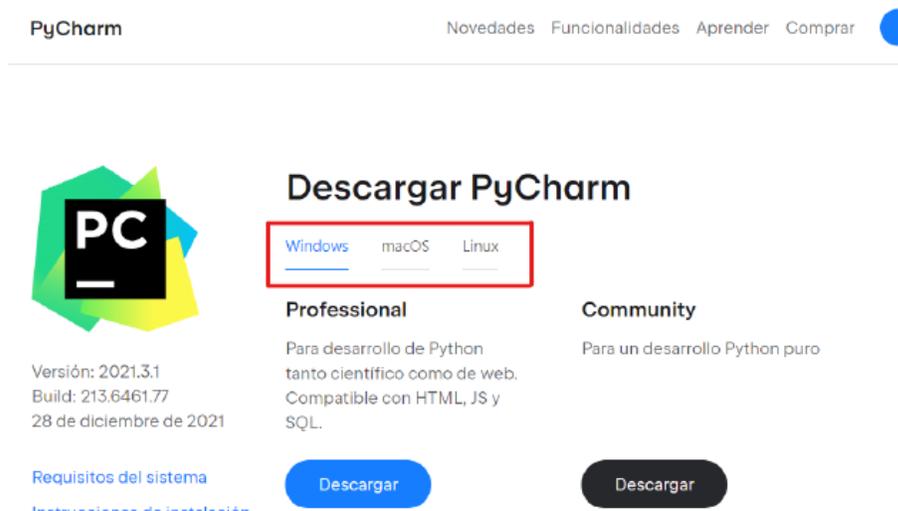
Manual de usuario

1. Requisitos de software para el manejo del programa

Para poder hacer uso de la presente interfaz de usuario es necesario tener instalado PyCharm Community Edition, el cual es un programa multiplataforma que puede ser ejecutado en Windows, macOS y Linux. Dentro de los requisitos del sistema se encuentran las versiones de 64 bits de Microsoft Windows 10 y 8, 2 GB de RAM como mínimo, 8 GB recomendados, 2.5 GB de espacio en disco duro, SSD recomendado, 1024x768 mínimo de resolución de pantalla, Python 2.7, Python 3.5 o más reciente.

PyCharm Community Edition se puede descargar directamente y de forma gratuita desde su página web: <https://www.jetbrains.com/es-es/pycharm/download/#section=windows>

Una vez allí se selecciona el sistema operativo y posteriormente la versión Community como se muestra marcado en rojo a continuación:



The screenshot shows the PyCharm website's download page. At the top, the 'PyCharm' logo is on the left, and navigation links for 'Novedades', 'Funcionalidades', 'Aprender', and 'Comprar' are on the right. Below the navigation bar, the PyCharm logo is displayed on the left, with version information: 'Versión: 2021.3.1', 'Build: 213.6461.77', and '28 de diciembre de 2021'. A link for 'Requisitos del sistema' is also present. The main heading is 'Descargar PyCharm'. Below this, there is a dropdown menu for selecting the operating system, with 'Windows' selected and highlighted by a red box. To the right of the dropdown, there are two columns for 'Professional' and 'Community' editions. The 'Professional' column describes it as suitable for Python development, including scientific and web, and compatible with HTML, JS, and SQL. The 'Community' column describes it as pure Python development. Each column has a 'Descargar' button.

PyCharm

Novedades Funcionalidades Aprender Comprar

Descargar PyCharm

Windows macOS Linux

Professional
Para desarrollo de Python tanto científico como de web. Compatible con HTML, JS y SQL.

Community
Para un desarrollo Python puro

Requisitos del sistema

Descargar Descargar



Versión: 2021.3.1
Build: 213.6461.77
28 de diciembre de 2021

[Requisitos del sistema](#)

[Instrucciones de instalación](#)

[Otras versiones](#)

[Software de terceros](#)

Descargar PyCharm

[Windows](#) [macOS](#) [Linux](#)

Professional

Para desarrollo de Python tanto científico como de web. Compatible con HTML, JS y SQL.

Descargar

[Prueba gratis](#)

Community

Para un desarrollo Python puro

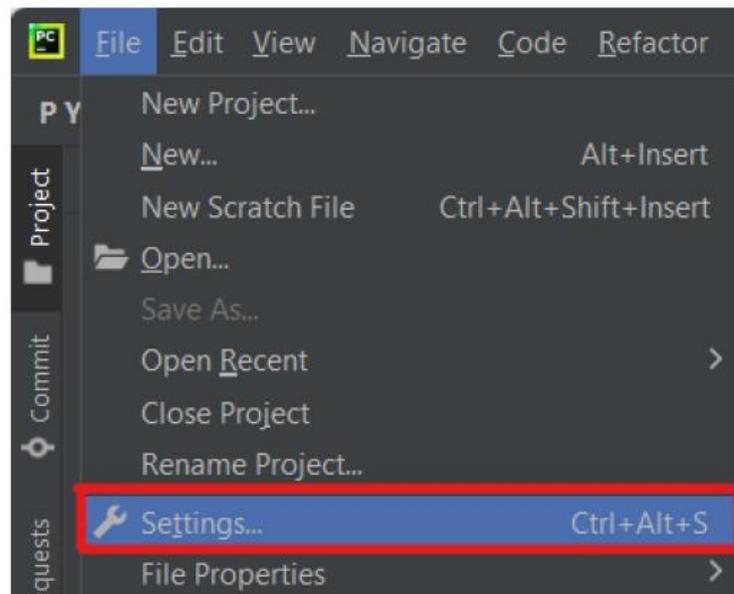
Descargar

Gratis, creado en código abierto

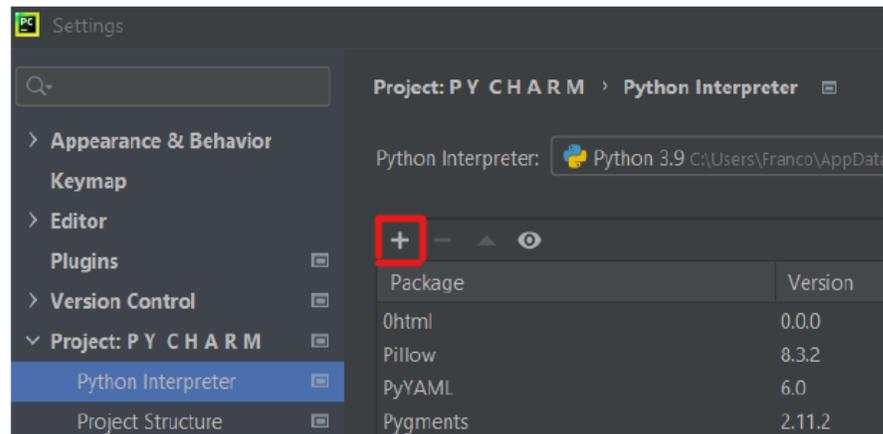
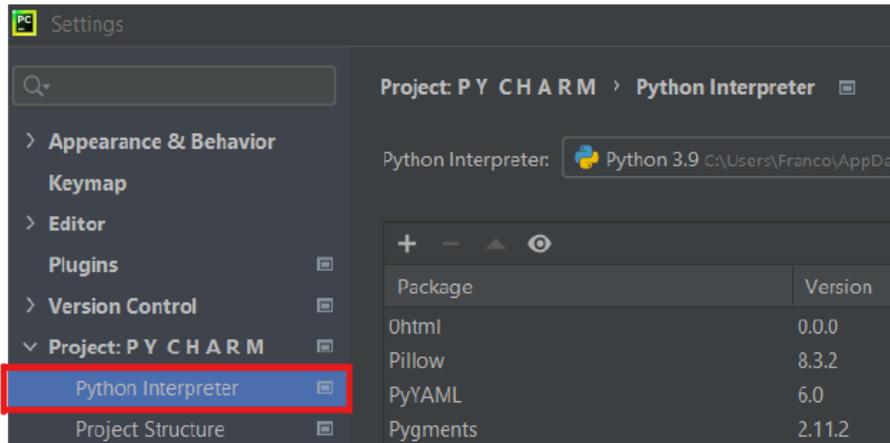
Una vez descargado el ejecutable se da click derecho, se selecciona correr como administrador y se siguen las indicaciones para su instalación.

2. Lenguajes y bibliotecas necesarias

Una vez abierto PyCharm Community Edition se va a la ruta File-Settings



A continuación, se selecciona la opción de Python Interpreter y se selecciona el lenguaje de programación con su respectiva versión, en este caso se utilizó Python 3.9 pero funciona con versiones anteriores. Una vez seleccionado el lenguaje y su versión en el signo + se agregan las bibliotecas necesarias. En este caso se utilizaron tkinter, pillow y matplotlib.



3. Descarga de la Interfaz

El programa con la interfaz gráfica fue subido a la plataforma github en donde se encuentra un archivo README.md que contiene las indicaciones necesarias para la descarga y ejecución de la interfaz, que a su vez está disponible públicamente en el siguiente enlace: <https://github.com/franodicandiaj/interfazrecursoeolico.git>

4. Menú principal

Una vez ejecutado se observa a continuación el Menú principal de la Interfaz.



Para observar los resultados de cada zona el usuario solo deberá pulsar los botones de ZONA 1: AEROPUERTO, ZONA 2: CEIBA, ZONA 3: TORCOROMA, ZONA 4: BELEN.

**Anexo 7. Certificación Platzi en el curso de ANÁLISIS DE
NEGOCIOS PARA CIENCIA DE DATOS**



Certifica a
GIAN FRANCO DI CANDIA JAIMES

Por participar y aprobar el
CURSO DE
**ANÁLISIS DE NEGOCIOS
PARA CIENCIA DE DATOS**



A handwritten signature in black ink, reading "Vander".

Christian Van Der Henst S
COO DE PLATZI



A handwritten signature in black ink, appearing to read "John Freddy Vega".

John Freddy Vega
CEO DE PLATZI

Certificación de aprobación online:

Aprobado el 6 de DICIEMBRE de 2021

10 horas de teoría y práctica

<https://platzi.com/@francodicandia888/>

Código: e699b546-1a7d-4003-8c9b-262f8e8f79e4

Anexo 8. Certificación Platzi en el curso de BÁSICO DE PYTHON



Certifica a

GIAN FRANCO DI CANDIA JAIMES

Por participar y aprobar el

CURSO
BÁSICO DE PYTHON



A handwritten signature in black ink, reading "Vander".

Christian Van Der Henst S
COO DE PLATZI



A handwritten signature in black ink, reading "John Freddy Vega".

John Freddy Vega
CEO DE PLATZI

Certificación de aprobación online:

Aprobado el 29 de DICIEMBRE de 2021

16 horas de teoría y práctica

<https://platzi.com/@francodicandia888/>

Código: c64ef518-3377-4574-9a7e-4fee4177a8b5

**Anexo 9. Certificación Platzi en el curso de entorno de trabajo
para CIENCIA DE DATOS CON JUPYTER NOTEBOOKS
Y ANACONDA**

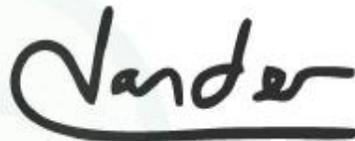


Certifica a

GIAN FRANCO DI CANDIA JAIMES

Por participar y aprobar el

CURSO DE ENTORNO DE TRABAJO PARA
**CIENCIA DE DATOS CON JUPYTER
NOTEBOOKS Y ANACONDA**

A handwritten signature in black ink that reads "Vander".

Christian Van Der Henst S
COO DE PLATZI

A handwritten signature in black ink that reads "John Freddy Vega".

John Freddy Vega
CEO DE PLATZI

Certificación de aprobación online:

Aprobado el 20 de ENERO de 2022

10 horas de teoría y práctica

<https://platzi.com/@francodicandia888/>

Código: 8bc748e8-15db-407d-a545-05c0718cceb