

SEP

TNM

INSTITUTO TECNOLÓGICO DE CULIACÁN



PRONÓSTICO DE LA VELOCIDAD DEL VIENTO
APLICANDO UN ESQUEMA DE VOTACIÓN

TESIS

PRESENTADA ANTE EL DEPARTAMENTO ACADÉMICO DE ESTUDIOS DE POSGRADO
DEL INSTITUTO TECNOLÓGICO DE CULIACÁN EN CUMPLIMIENTO PARCIAL DE LOS
REQUISITOS PARA OBTENER EL GRADO DE

MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

POR:

MANJARREZ MONTELONGO GIOVANNI
INGENIERO EN SISTEMAS COMPUTACIONALES

DIRECTOR DE TESIS:

DR. RODRÍGUEZ RANGEL HÉCTOR

CULIACÁN, SINALOA

AGOSTO, 2018

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

Culiacán, Sin., 27 de Agosto del 2018

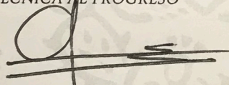
OFICIO: DEPI-339/VIII/2018
ASUNTO: **Autorización Impresión**

ING. GIOVANNI MANJARREZ MONTELONGO
ESTUDIANTE DE LA MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN
PRESENTE.


Por medio de la presente y en virtud de que ha completado los requisitos para el examen de grado de la **Maestría en Ciencias de la Computación**, se concede autorización para la impresión de la tesis titulada: **"PRONÓSTICO DE LA VELOCIDAD DEL VIENTO APLICANDO UN ESQUEMA DE VOTACIÓN"**, bajo la dirección del(a) **Dr. Héctor Rodríguez Rangel**

Sin otro particular reciba un cordial saludo.

ATENTAMENTE
"CON LA TÉCNICA AL PROGRESO"



M.C. MARÍA ARACELY MARTÍNEZ AMAYA
JEFE(A) DE LA DIVISIÓN DE ESTUDIOS DE
POSGRADO E INVESTIGACIÓN



SEP TecNM
Instituto Tecnológico
de Culiacán
División de Estudios
de Posgrado e Investigación

C.c.p. archivo

MAMA/lucy *

“PRONÓSTICO DE LA VELOCIDAD DEL VIENTO APLICANDO UN ESQUEMA DE VOTACIÓN”

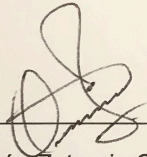
Tesis presentada por:

ING. GIOVANNI MANJARREZ MONTELONGO

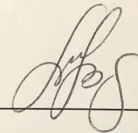
Aprobada en contenido y estilo por:



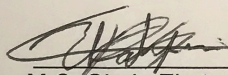
Dr. Héctor Rodríguez Rangel
Director de Tesis



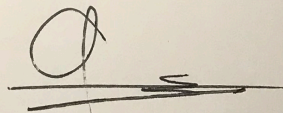
Dr. Ramón Zatarain Cabada
Secretario



Dra. María Lucía Barrón Estrada
Vocal -1



M.C. Gloria Ekaterine Peralta Peñuñuri
Vocal -2



M.C. María Aracely Martínez Amaya
Jefe(a) de la División de Estudios de
Posgrado e Investigación

Dedicatoria

Esta tesis es dedicada a mis padres Adán Manjarrez e Imelda Montelongo, sin sus consejos y apoyo incondicional no habría sido posible lograr este paso importante en mi carrera profesional. Además, dedico este trabajo a mis compañeros de posgrado, ya que durante la convivencia que tuve en este periodo, se obtuvo una buena relación de trabajo lo cual generó una buena amistad entre todos. También dedico mi trabajo a mi novia María Muñiz por apoyarme en todo momento durante el tiempo comprendido de los estudios de maestría. Y finalmente quiero dedicar este trabajo a mi mismo por ser una persona que lucha por sus objetivos y no se rinde a obstáculos que se encuentren en el camino del desarrollo profesional.

Agradecimientos

Quisiera reconocer y agradecer al Consejo Nacional de Ciencia y Tecnología (CONACyT) por haberme brindado el apoyo económicamente para llevar a cabo mis estudios de maestría, además de darme la oportunidad de realizar dos estancias de investigación a nivel nacional e internacional. Así como al Instituto Tecnológico de Culiacán por abrirme las puertas y brindarme la oportunidad de cursar y obtener el grado de maestría en ciencias de la computación.

Además, agradezco al Dr. Héctor Rodríguez Rangel por ser mi guía a lo largo de este camino y brindarme su apoyo y conocimientos que sin ellos no hubiera sido posible el desarrollo de este trabajo. También agradecer a los demás doctores de posgrado Lucía Barrón, Ramón Zatarain, Ricardo Quintero y Luis Carlos Santillán, ya que fueron parte fundamental en mi desarrollo profesional durante el periodo comprendido de los estudios de maestría.

Reconozco a los doctores Juan José Flores y Vicenç Puig Cayuela por compartir sus conocimientos y brindarme la ayuda posible para realizar el trabajo de tesis durante la estancia de investigación en la Universidad Michoacana de San Nicolás de Hidalgo en la ciudad de Morelia, Michoacán y en la Universidad Politécnica de Cataluña de la ciudad de Barcelona, España.

Finalmente quiero agradecer a la coordinadora del departamento de posgrado de sistemas M.C. Gloria Ekaterine Peralta y a la secretaria Lucy por toda la ayuda y disposición que me brindaron ante cualquier actividad que necesitara durante el tiempo comprendido de estudios de maestría.

Declaración de autenticidad

Por la presente declaro que, salvo cuando se haga referencia específica al trabajo de otras personas, el contenido de esta tesis es original y no se ha presentado total o parcialmente para su consideración para cualquier otro título o grado en esta o cualquier otra Universidad. Esta tesis es resultado de mi propio esfuerzo, trabajo, dedicación y no incluye nada que sea resultado de algún trabajo realizado en colaboración, salvo que se indique específicamente en este documento.

Giovanni Manjarrez Montelongo. Culiacán, Sinaloa, México, 2018

Resumen

La energía ha sido de gran importancia en las actividades del hombre y mayormente es producida por medios fósiles, la cual, ocasiona daños irreversibles al planeta debido a grandes cantidades de dióxido de carbono (CO₂) emite a la atmósfera. Esta energía abarca más de dos tercios de la producción, según estudio realizado por Observ'ER & Foundation Énergies pour le Monde en 2012.

Se ha buscado un mecanismo para disminuir el daño ambiental generado por este tipo de energía (no renovable); una alternativa de solución es la energía renovable. La energía eólica es un tipo de energía renovable y ha presentado un crecimiento anual entre el 2002 y 2012 de un 26.1%. Con el crecimiento de esta energía (limpia) se han encontrado desafíos para satisfacer la demanda de producción; una necesidad principal es tener una estimación de la cantidad de energía que se producirá en un determinado tiempo. Esta estimación se genera realizando predicciones de variables involucradas en la producción de energía.

La propuesta de este trabajo se enfoca en realizar predicciones de series de tiempo de la velocidad del viento aplicando diferentes técnicas de inteligencia artificial (IA), con la finalidad de optimizar la producción de la energía eólica. Con el conjunto de predicciones se implementa un esquema de fusión (EF); la fusión se realiza utilizando diferentes enfoques (media y mediana aritmética, y agrupamiento de datos). Con esta implementación de fusión se procede a seleccionar la mejor fusión mediante la comparación de resultados obtenidos tomando como medida de aptitud el error cuadrático medio (MSE).

Para probar la metodología propuesta se utilizaron series de tiempo recolectadas de diferentes localidades del estado de Michoacán. Se observó que el MSE obtenido por el EF fue menor en comparación al MSE obtenido por métodos de IA. Esto se refiere a que el pronóstico generado por el EF fue más preciso.

Palabras clave

Energía limpia.

Series de Tiempo.

Pronóstico.

Inteligencia Artificial.

Redes Neuronales.

Vecinos Cercanos.

Regresión Lineal.

Árboles Aleatorios.

Fusión de Datos

Mediana Aritmética.

Media Aritmética.

Agrupamiento de Datos.

Índice general

Índice de figuras	X
Índice de tablas	XII
1. Introducción	1
1.1. Definición del problema	2
1.2. Objetivo principal	3
1.2.1. Objetivos específicos	3
1.3. Justificación	4
1.4. Estructura del documento	5
2. Marco teórico	6
2.1. Energía	6
2.1.1. Energía no renovable	7
2.1.2. Energía renovable	8
2.1.3. Energía eólica	8
2.2. Series de tiempo	9
2.2.1. Pronóstico de series de tiempo	10
2.3. Métodos de predicción	11
2.3.1. Métodos estadísticos	12
2.3.2. Métodos de inteligencia artificial	13
2.3.2.1. Redes neuronales artificiales	13
2.3.2.2. k - vecinos cercanos	15
2.3.2.3. Regresión lineal	15
2.3.2.4. Árboles aleatorios	17
2.3.3. Métodos híbridos	18
2.4. Fusión de Datos	18
2.4.1. Mediana Aritmética	19
2.4.2. Media Aritmética (Promedio)	19
2.4.3. Agrupamiento de Datos	20
3. Estado del arte	22
3.1. Pronóstico	22
3.2. Métodos de predicción con IA	23
3.3. Fusión de datos	26

4. Desarrollo	28
4.1. Creación de la Base de Datos	29
4.1.1. Procesamiento de los datos	29
4.1.1.1. Normalización	30
4.1.2. Generación de la base de datos	31
4.2. Construcción de modelos de predicción	32
4.2.1. Pronóstico con aplicando redes neuronales	33
4.2.2. Pronóstico con k -vecinos cercanos	35
4.2.3. Pronóstico con regresión lineal	36
4.2.4. Pronóstico con árboles aleatorios	38
4.3. Esquema de fusión	39
4.3.1. Implementación de la mediana	39
4.3.2. Implementación de la promedio	39
4.3.3. Implementación de agrupamiento de datos	41
4.4. Selección de la mejor fusión	42
5. Análisis de los Resultados	44
5.1. Resultados	44
6. Conclusiones	49
6.1. Conclusiones	49
6.2. Aportaciones	50
6.3. Trabajo futuro	50

Índice de figuras

2.1.	Imagen representativa de una serie de tiempo, fuente: (Montgomery et al., 2015).	10
2.2.	Imagen representativa de una neurona biológica, fuente: (Negnevitsky, 2005).	13
2.3.	Imagen que representa un ejemplo de la arquitectura de una red neuronal artificial.	14
2.4.	Ejemplo de clasificación basada en la similaridad de los individuos k -NN	16
2.5.	Imagen representativa del pronóstico aplicando regresión lineal.	16
2.6.	Imagen representativa de la implementación de árboles aleatorios en pronósticos.	17
2.7.	Ilustración del rango permitido para calcular el promedio de datos para el pronóstico de series de tiempo.	21
2.8.	Muestra de elementos agrupados de un conjunto de datos.	21
4.1.	Proceso del pronóstico de la velocidad del viento.	29
4.2.	Muestra del tamaño de ventana de una serie de tiempo.	30
4.3.	Proceso de normalización de las series de tiempo.	31
4.4.	Proceso del pronóstico de la velocidad del viento definido para cada técnica de predicción.	33
4.5.	Implementación del esquema de fusión con los modelos de predicción utilizados en este trabajo.	34
4.6.	Arquitectura de ANN con m observaciones pasadas, h neuronas en su capa oculta y una única salida (y_{t+1}).	34
4.7.	Datos bidimensionales esparcidos gráficamente.	37
4.8.	Predicción de la serie de tiempo aplicando regresión lineal.	37
4.9.	Proceso del pronóstico de series de tiempo de la velocidad del viento aplicando árboles aleatorios.	38
4.10.	Proceso de la aplicación de la mediana con los diferentes pronosticadores.	40
4.11.	Proceso de la aplicación del promedio con los diferentes pronosticadores.	40
4.12.	Valores de la serie de tiempo pronosticados con las técnicas de IA.	41
4.13.	Proceso de cálculo de agrupamiento de datos para obtener el pronóstico de la serie de tiempo.	42
5.1.	Resultados de los pronósticos implementando un esquema de votación de series de tiempo tomadas de estaciones eólicas de diferentes localidades del Michoacán.	47

5.2. Resultados de los pronósticos implementando un esquema de votación de series de tiempo tomadas de estaciones eólicas de diferentes localidades del Michoacán. 48

Índice de tablas

4.1. Base de datos de la serie de tiempo.	32
5.1. Series de tiempo con características caóticas	45
5.2. Promedio con los diferentes métodos de IA	45
5.3. Resultados de los pronósticos en cada serie de tiempo	46
5.4. Pronóstico con mejor resultado en cada serie de tiempo.	46

Capítulo 1

Introducción

Actualmente, la energía es un factor de gran importancia en las actividades del hombre y producida por medios fósiles, la cual, ocasiona daños irreversibles al planeta. Es por esto que la producción de energía se ha convertido en uno de los principales cuestionamientos tanto económicos, como medio ambientales. Además, la demanda de energía ha ido creciendo en los últimos años, dando como resultado daños ambientales al momento de producirla. Esta energía producida por medios fósiles se denomina energía no renovable (Observ, 2012).

Para reducir la problemática generada por la energía no renovable y mejorar la calidad de vida en el planeta se encuentra la alternativa de la energía limpia ó energía producida por medios renovables (radiación solar, viento, calor, movimiento, etc.). La energía generada por la radiación solar y el viento han mostrado un global crecimiento anual entre el año 2002 y 2012 con un 50 % y 26.1 % respectivamente (Observ, 2012). Esta energía limpia busca contrarrestar los daños ambientales que genera la producción de la energía no renovable.

Con el crecimiento de la energía renovable, se han encontrado muchos desafíos para satisfacer la demanda de producción de energía con el objetivo de disminuir el costo de producción, mejorar la planificación de las plantas energéticas, etc. En diferentes trabajos se ha mencionado que una de las necesidades principales para enfrentar estos desafíos es realizar pronóstico de variables involucradas en la producción de energía, es decir, tener una estimación de la cantidad de energía que se producirá en los próximos minutos, horas, días o meses. Estas variables involucradas que generan energía limpia son recolectadas por medio de la naturaleza: velocidad del viento, temperatura, humedad, presión atmosférica, ó niveles de presa, etc.

Este trabajo específicamente se enfoca en realizar predicciones de la velocidad del viento con el objetivo de resolver el problema de la estimación de energía que se producirá, en este caso el viento. En la literatura se encuentra diferentes enfoques utilizados para realizar la tarea de predicción. (Chang, 2014) menciona que existen diferentes métodos de pronóstico como: estadístico, de persistencia, físico, inteligencia artificial y un enfoque híbrido el cual consiste en una combinación de dos o más métodos de predicción.

De acuerdo con (Rodríguez et al., 2017, Yuan-Kang et al., 2010 y Chang, 2013), se ha comprobado que los métodos de inteligencia artificial (IA) tienen mejores resultados con respecto a otros métodos de predicción. Mediante la comparación de resultados obtenidos con distintos métodos de predicción. Los métodos de IA que sobresalen son: redes neuronales, vecinos cercanos, regresión lineal, entre otros. Este trabajo pretende explorar las diferentes técnicas de IA para realizar pronóstico de series de tiempo; además, con el resultado de esta exploración, se realiza un esquema de fusión, lo cual permite mejorar el pronóstico obtenido mediante un simple pronosticador.

1.1. Definición del problema

De acuerdo con el censo realizado por Observ'ER & Foundation Énergies pour le Monde en 2012, la energía producida por medios fósiles abarca más de dos tercios de la producción total de energía (68.1 %) (Observ, 2012). Esta energía ocasiona contaminación hacia el medio ambiente, la cual, no puede ser resuelta con facilidad debido a las grandes cantidades de dióxido de carbono (CO₂) que genera en la atmósfera, provocando el aumento de gases de efecto invernadero.

Otra problemática de la producción de energía es el residuo radioactivo. El residuo radiactivo es aquel que contienen elementos químicos radiactivos y que no tienen un propósito práctico. Este residuo también puede generarse durante el proceso de combustible para los reactores, armas nucleares o en las aplicaciones médicas como la radioterapia o la medicina nuclear.

Para reducir el daño ambiental se opta por utilizar la energía renovable; esta energía es abundante en todo el planeta; se dice que ocho países tienen el 81 % de reserva de petróleo

crudo, seis países tienen el 70% de reserva de gas natural y ocho países el 89% reservas de carbón (Boyle, 1996). La energía eólica es una de la más utilizada ya que proviene del recurso natural el viento; es una fuente de energía limpia abundante en la mayor parte del mundo, de bajo costo, sostenible, segura y popular (Boyle, 1996). Para obtener la cantidad necesaria de energía producida por el viento y satisfacer la demanda de producción es necesario realizar predicciones de las variables involucradas, en este caso la velocidad del viento.

Para almacenar la información que se pronosticará en un determinado tiempo, se utilizan series de tiempo. Las series de tiempo de la velocidad del viento tienen un comportamiento caótico. Este comportamiento indica que es sensible a cambios, es decir, son difíciles de predecir (Montgomery et al., 2015). Además, la velocidad del viento es irregular e intermitente, esto genera que en la mayoría de los casos las series de tiempo presenten datos incompletos, por razones como: falla en sensores, problemas de comunicación, errores humanos, etc. (Hayashi and Kermanshahi, 2001). Con este desafío de las series de tiempo caóticas se realizan diferentes modelos de predicción.

1.2. Objetivo principal

Implementar un modelo predictivo de la velocidad del viento mediante una fusión de datos provenientes de diferentes pronosticadores con la finalidad de mejorar los métodos de pronóstico simples de IA.

1.2.1. Objetivos específicos

- Diseñar un modelo de predicción implementando diferentes métodos de inteligencia artificial.
- Implementar los diferentes métodos de IA para realizar el pronóstico de la velocidad del viento.
- Realizar una fusión de datos (pronósticos) obtenidos con la finalidad de mejorar la predicción de series de tiempo de la velocidad del viento.

- Analizar los resultados obtenidos al fusionar los datos pronosticados con cada método de IA.

1.3. Justificación

El uso de la energía renovable permite reducir de manera significativa la producción de energía proveniente del uso de combustibles fósiles y nucleares, reduciendo así los impactos ambientales asociados. La mayoría de las fuentes de energía renovable permiten evitar las formas de daño ambiental (Boyle, 1996).

El daño de la producción de energía hacia el medio ambiente ha requerido la necesidad de realizar predicciones sobre la producción de la energía renovable; se ha generado un incremento de enorme potencial en recursos renovables. De acuerdo con (Hong et al., 2014), los sistemas de energía eléctrica son los más complejos hechos por el hombre, produciendo y suministrando electricidad a 5.6 billones de personas. La industria de la energía eléctrica necesita pronósticos de suministro, demanda, precio y energía para planificar y operar la red.

La producción de energía renovable tiene un beneficio que impacta en el mejoramiento del medio ambiente. El sistema de energía de Estados Unidos ha hecho una estimación para que en el año 2030, un pronóstico será valorado en más de 3 billones de dólares por año (Piwko and Jordan, 2010). Desde el año 2004 las fuentes renovables (solar, eólica, biomasa, etc.) se han desarrollado en un 3.2%, gracias a los resultados obtenidos en diferentes áreas del medio ambiente. Con la continuación del desarrollo de predicción que impacten en el desarrollo de la energía, la reducción del daño ambiental será de mayor peso.

La capacidad de pronosticar la velocidad del viento es esencial para la buena integración de la energía eólica en los sistemas eléctricos. De acuerdo con (Barber et al., 2010), el pronóstico de la velocidad del viento para la industria de energía eólica ayuda a disminuir los costos de producción en el área de planeación, etc. Además de incentivar al sector público y privado a invertir en la generación de energía por medio del viento.

1.4. Estructura del documento

Este trabajo está organizado de la siguiente manera: Capítulo 2 describe el marco teórico donde se especifican conceptos básicos acerca de la producción de energía, series de tiempo, pronósticos, metodologías de inteligencia artificial como redes neuronales, vecinos cercanos, regresión lineal y árboles aleatorios. Además, el concepto de fusión de datos donde se mencionan diferentes opciones de aplicación del esquema de fusión. En el Capítulo 3 se muestran trabajos relacionados con hacer predicciones, implementando metodologías estadísticas, de inteligencia artificial y fusión de datos. En el Capítulo 4 se presenta el desarrollo del diseño de la metodología propuesta, mostrando los pasos y algoritmos de ejecución de la predicción de series de tiempo de la velocidad del viento. El Capítulo 5 muestran los resultados obtenidos de la metodología y la comparación de los resultados con la aplicación del esquema de fusión. Finalmente en el Capítulo 6 se presentan las conclusiones sobre el trabajo realizado.

Capítulo 2

Marco teórico

Realizar pronóstico de la velocidad del viento no es una tarea sencilla, es necesario tener un conjunto de datos históricos para realizar las estimaciones. Estos datos se almacenan en lo que se denominan como series de tiempo. Las predicciones se realizan desde diferentes enfoques como: persistencia, estadísticos o de inteligencia artificial, en este trabajo se utilizó el enfoque de la inteligencia artificial. En este capítulo se presentará una breve descripción de conceptos involucrados en el trabajo como: energía, energía eólica, series de tiempo, pronóstico, técnicas de predicción y métodos de inteligencia artificial implementados en el trabajo.

2.1. Energía

La energía se denomina como la capacidad de producir un efecto, no necesariamente desde un inicio de conservación, ya que, es una propiedad física que puede transformarse y que interviene en procesos de cambios de estado (Alomá and Malaver, 2007). Además, la energía es una actividad humana; es vista como un combustible. Está relacionada con el movimiento o con alguna actividad visible (Doménech et al., 2001).

Existen diferentes formas que se puede manifestar la energía y que puede sufrir transformaciones, entre ellas se encuentra la energía química, la cual se encuentra presente en los combustibles fósiles (carbón, petróleo y gas natural), madera y otras sustancias combustibles. Cuando un combustible arde en el aire, la energía química del combustible se transforma en otra forma de energía que suele denominarse energía térmica (Doménech et al., 2001).

De acuerdo con (Alomá and Malaver, 2007), la energía es todo lo que posee un cuerpo y que realiza una acción; todo cuerpo que tiene un movimiento tiene energía. Puede almacenarse dentro de objetos, un ejemplo de ello es el carbón, ya que tiene energía almacenada, en este caso la energía es una sustancia.

De acuerdo con (Pérez-Landazábal et al., 1995), la energía es una magnitud abstracta que está ligada al estado dinámico de un sistema cerrado y que permanece invariable con el tiempo. Se trata de una abstracción que se le asigna al estado de un sistema físico. Debido a diversas propiedades, todos los cuerpos poseen energía.

En el campo de la física, se llega a determinar que se produce la mención a diversos tipos de energía, mecánica, cinética, eólica, entre otras (Observ, 2012). En el medio ambiente existen recursos que son capaces de suministrar y brindar energía en cualquier forma, por lo que se consideran fuentes naturales de energía o recursos energéticos.

2.1.1. Energía no renovable

La energía no renovable es el tipo de energía que se obtiene de un recurso natural limitado, por lo que una vez que éste se agote, no podrá sustituirse. Como no existe un sistema de producción o de extracción viable, cuando se agoten las reservas de los recursos que permiten generar una energía no renovable, la misma dejará de existir.

Se llama energía no renovable a aquella que proviene de fuentes agotables, como la procedente del petróleo, el carbón o el gas natural. En cambio, la energía renovable es virtualmente infinita, como la eólica (generada por la acción del viento) y la solar.

La energía no renovable procede de combustibles fósiles o combustibles nucleares. En el primer grupo se encuentra el petróleo, carbón y gas natural. Estos combustibles son fáciles de extraer, pero su reserva se agotan en el corto o mediano plazo. Una vez que no existan más reservas de petróleo, por ejemplo, el ser humano no estará en condiciones de crear nuevas (Cheng, 2017).

2.1.2. Energía renovable

La energía renovable es un recurso natural que se aprovecha industrialmente a partir de la aplicación de tecnología y de diversos recursos asociados. La noción de energía renovable hace mención al tipo de energía que se obtiene de fuentes naturales virtualmente inagotables, ya que contiene una inmensa cantidad de energía (Lei et al., 2009).

La energía renovable se obtiene de corrientes continuas o repetitivas de energía que se repiten en el entorno natural. La mayoría de las fuentes de energía renovable se derivan de la radiación solar, incluido el uso directo de energía solar para generación de electricidad, y formas indirectas como energía del viento, las olas y el agua corriente, y de plantas y animales (madera, paja, estiércol y otros desechos de plantas) (Boyle, 1996).

Las fuentes de energía de las mareas son el resultado de la atracción gravitatoria de la luna y el sol, y la energía geotérmica proviene del calor generado en la tierra. La energía proveniente de desechos de todo tipo también se incluye a menudo bajo el título de fuentes renovables. Por otro lado, en su desarrollo tecnológico, las fuentes de energía renovable van desde tecnologías bien establecidas y maduras hasta aquellas que requieren mayor investigación y desarrollo (Boyle, 1996).

La energía renovable ofrece al planeta la oportunidad de reducir las emisiones de carbono, limpiar el aire y poner a nuestra civilización en una base más sostenible. Las fuentes de energía renovable son una parte esencial de una estrategia general de desarrollo sostenible. Ayudan a reducir la dependencia de las importaciones de energía, lo que garantiza un suministro sostenible y la protección del clima (Behl et al., 2012).

Además, las fuentes de energía renovable pueden ayudar a mejorar la competitividad de las industrias a largo plazo y tener un impacto positivo en el desarrollo regional y el empleo. La energía renovable proporcionará un grupo de fuentes de energía más diversificado, equilibrado y estable (Behl et al., 2012).

2.1.3. Energía eólica

La energía eólica es ejemplo de energía renovable debido a que su utilización supone una mínima huella ambiental. El término eólico es un adjetivo que es relativo al viento; se conoce

como viento a la corriente de aire que se produce naturalmente en la atmósfera. La energía eólica proviene del viento y trata de un tipo de energía cinética producida por el efecto de las corrientes de aire (López, 2012).

La energía eólica ha realizado aportes importantes en el aspecto técnico como económico, ya que, ha mejorado aspectos como la gestión y mantenimiento de parques eólicos, la integración de la energía eléctrica en la red, la versatilidad y adaptación del diseño de aerogeneradores, etc. (López, 2012).

La forma típica de aprovechar esta energía eólica es a través de la utilización de aerogeneradores o turbinas de viento; para obtener electricidad el movimiento de los aerogeneradores o aspas eólicas acciona un generador eléctrico que convierte la energía mecánica de la rotación en energía eléctrica. El funcionamiento es bastante simple y lo que se va dificultando es la construcción de aerogeneradores que sean cada vez más eficientes.

La producción de energía eólica depende del calentamiento de la superficie terrestre por la radiación solar, lo que provoca viento. Este viento es transformado por aeromotores, máquinas eólicas, aerogeneradores y turbinas eólicas que convierten la energía cinética del viento en energía mecánica para que a su vez esta produzca electricidad (Moragues and Rapallini, 2004).

2.2. Series de tiempo

Una serie de tiempo se define como un conjunto de observaciones cuantitativas, arregladas en orden cronológico, donde generalmente se asume que el tiempo es una variable discreta (Kirchgässner and Wolters, 2007). Las series de tiempo se encuentran en los campos de ingeniería, ciencia, economía, entre otras (Brockwell and Davis, 2016).

Una serie de tiempo se representa en una gráfica de secuencia; esto es, representar cada observación x frente a una constante de tiempo t . La Figura 2.1 muestra como evoluciona una serie a lo largo del tiempo. En una serie de tiempo se presentan diferentes características: presencia de tendencia (comportamiento a lo largo del tiempo), estacionalidad (comportamiento periódico) y heterocedasticidad.

El análisis de series de tiempo es el procedimiento por el cual se identifican y aíslan

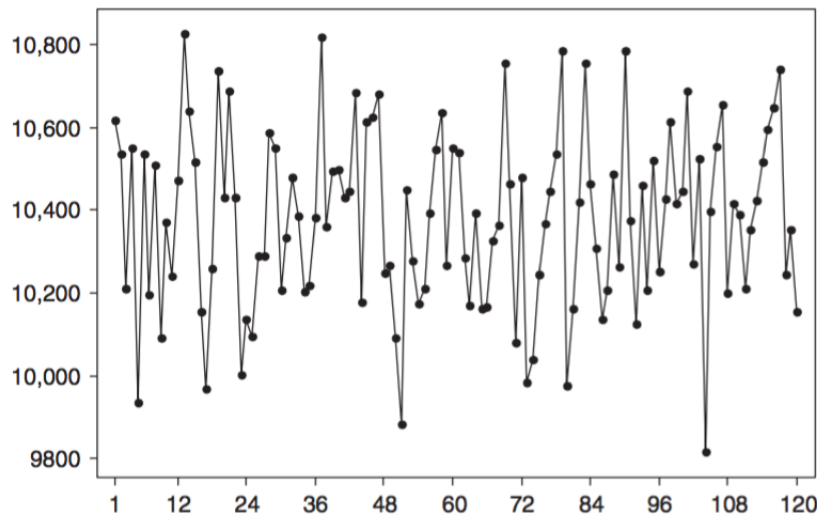


Figura 2.1: Imagen representativa de una serie de tiempo, fuente: (Montgomery et al., 2015).

los factores relacionados con el tiempo que influyen en los valores observados en las series de tiempo. Una vez identificado estos factores, se contribuye a la interpretación de valores históricos de series de tiempo. Las gráficas de series de tiempo pueden revelar patrones tales como aleatoriedad, tendencia, cambios de nivel, períodos o ciclos, observaciones inusuales o una combinación de patrones (Montgomery et al., 2015).

2.2.1. Pronóstico de series de tiempo

Pronosticar se refiere aquel conocimiento anticipado de lo que sucederá en un determinado futuro, mediante diferentes indicios, señales u observaciones previas acerca de un problema o tarea en particular (Montgomery et al., 2015). Generalmente la tarea de pronosticar se enfoca en la optimización en procesos en diferentes áreas como los niveles de inventario, la capacidad de producción o los niveles de personal. La tarea de pronosticar debe ser tan precisa como sea posible, dado que esta ligada a actividades de toma de decisiones, ventas de productos, índices de la bolsa de valores, demanda eléctrica, etc.

Existen dos variables estructurales principales que definen un pronóstico de serie de tiempo: el período, que representa el nivel de agregación, los más comunes son meses, semanas y días en la cadena de suministro (para la optimización del inventario). Y el horizonte, que representa la cantidad de períodos por adelantado que deben ser pronosticados. En la cadena

de suministro, el horizonte es generalmente igual o mayor que el tiempo de entrega.

El término de predicción se refiere a la estimación de series temporales, o datos instantáneos; la tarea de pronosticar está enfocada en el plan de demanda diaria de negocios. La práctica del plan de demanda también se refiere al pronóstico de la cadena de suministros. El pronóstico crítico y continuo se necesita para obtener un buen resultado durante la planificación de un proyecto (Rodríguez et al., 2017).

La tarea de pronosticar se efectúa en diferentes escalas de tiempo. La clasificación a escalas de tiempo de los métodos de predicción de la energía eólica es diferente en varias descripciones que se encuentran en la literatura. De acuerdo con (Chang, 2014), las escalas de tiempo los métodos de predicción de la energía eólica las define en 4 categorías: 1) Pronóstico a muy corto plazo: Se efectúa en minutos a una hora de anticipación. 2) Pronóstico a corto plazo: Es de una hora a varias horas por delante. 3) Pronóstico a mediano plazo: De varias horas a una semana de anticipación. 4) Pronóstico a largo plazo: De una semana a un año o más por delante.

La ventaja de realizar pronósticos es apoyar al toma de decisiones en distintas actividades del hombre, el caso de la velocidad del viento beneficia en la reducción del costos de producción y planeación de plantas eólicas. Además las predicciones permiten mayor flexibilidad en la elaboración de los planes, sobre todo cuando se sirve de los métodos de múltiples escenarios (Rodríguez et al., 2017).

De acuerdo con (Barber et al., 2010), la importancia de pronosticar el viento para la industria de energía eólica se deriva de tres hechos: 1) La potencia agregada producida y consumida en todo el sistema eléctrico debe estar en equilibrio para garantizar la fiabilidad y la seguridad del suministro eléctrico. 2) La potencia de un parque eólico es muy variable ya que depende fuertemente de la velocidad y dirección del viento. 3) No existe un mecanismo eficiente y rentable de almacenamiento de energía eléctrica.

2.3. Métodos de predicción

Los pronósticos de energía eólica siguen desempeñando un papel clave para abordar los desafíos operacionales en el suministro eléctrico. Recientemente, se han empleado varios

métodos para la predicción de la energía eólica (Wu and Hong, 2007, Lei et al., 2009 y Yadav and Sahu, 2017). De acuerdo con (Chang, 2014), los métodos de predicción se clasifican en persistencia, físicos, estadísticos, correlación espacial y de inteligencia artificial.

Las técnicas tradicionales de predicción se basan en modelos lineales, sin embargo, muchas de las series de tiempo que se encontraron en la práctica, exhiben características no presentes en procesos lineales. Entonces, estas series de tiempo muestran características dinámicas enriquecidas, incluyendo sensibilidad a condiciones iniciales, conocidas como comportamiento caótico. Este comportamiento hace que la complejidad de predicción se incremente (Leven and Koch, 1981).

En el pasado se han propuesto varios métodos para pronosticar series de tiempo caóticas. Primero fue el uso de modelos no lineales (NLM, por sus siglas en inglés) (Farmer and Sidorowich, 1987), sin embargo, los principales problemas de los métodos de predicción era la selección adecuada de modelos y la dependencia de datos. Otros métodos de predicción que se utilizan para resolver el problema de predicción de series de tiempo son las técnicas de inteligencias artificial, entre ellas, se encuentran las redes neuronales artificiales, vecinos cercanos, lógica difusa, entre otros (Maguire et al., 1998).

2.3.1. Métodos estadísticos

Entre los diferentes métodos de predicción se encuentran los estadísticos; estos métodos proporcionan modelos de predicción relativamente económicos que no requieren de ningún dato que va más allá de datos históricos (Lei et al., 2009).

En estos modelos estadísticos, se utilizan los datos históricos de sistemas de conversión de energía del viento (por su siglas en inglés *WECS*). Según (Chang, 2014), los modelos estadísticos son fáciles de modelar en comparación con otros modelos. Básicamente, los métodos estadísticos son buenos para períodos de tiempo cortos. La desventaja de estos métodos es que el error de predicción aumenta a medida que aumenta el tiempo de predicción.

Entre los métodos estadísticos existen diferentes técnicas de aplicación: auto regresivo, (AR, por sus siglas en inglés) el auto promedio móvil regresivo (ARMA), el auto promedio móvil integrado regresivo (ARIMA), el enfoque bayesiano y las predicciones grises. Los métodos estadísticos son utilizados para resolver los problemas de ingeniería, economía y

ciencias naturales que tienen una gran cantidad de datos donde las observaciones son interdependientes.

2.3.2. Métodos de inteligencia artificial

Además de los métodos estadísticos, se encuentra los métodos de inteligencia artificial (IA) donde destacan las técnicas de redes neuronales, vecinos cercanos, regresión lineal, árboles aleatorios, etc.

2.3.2.1. Redes neuronales artificiales

Las redes neuronales artificiales (ANN, por sus siglas en inglés) simulan ciertas características propias de los humanos, como la capacidad de memorizar y de asociar hechos. Una red neuronal se define como un modelo de razonamiento basado en el cerebro humano. El ser humano es capaz de enfrentarse a situaciones y las resuelve acudiendo a la experiencia acumulada. Así, parece claro que una forma de aproximarse al problema es en la construcción de sistemas que sean capaces de reproducir esta característica humana (Negnevitsky, 2005).

Las ANN son técnicas de IA compuestas por un gran número de elementos simples, elementos de procesos interconectados. Los cuales procesan información por medio de su estado dinámico como respuesta a entradas externas. La Figura 2.2 representa una red neuronal biológica.

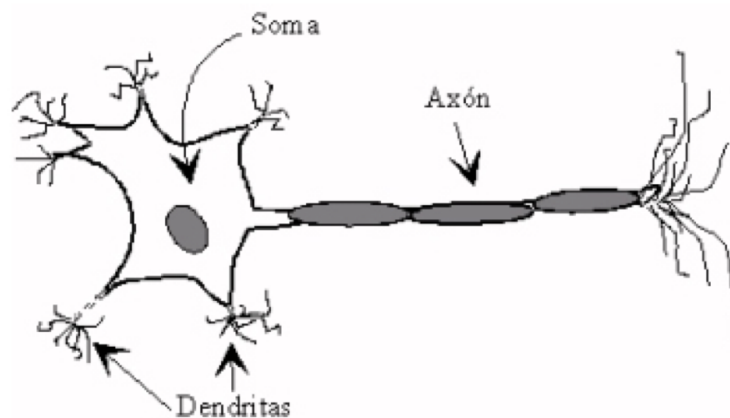


Figura 2.2: Imagen representativa de una neurona biológica, fuente: (Negnevitsky, 2005).

Las redes neuronales no son más que un modelo artificial y simplificado del cerebro humano. Son el ejemplo perfecto disponible para un sistema que es capaz de adquirir conocimiento a través de la experiencia. Una red neuronal es un sistema para el tratamiento de la información, cuya unidad básica de procesamiento está inspirada en la célula fundamental del sistema nervioso humano: la neurona.

Las ANN tratan problemas no lineales y complejos en términos de clasificación, regresión, o pronóstico. De acuerdo con (Yuan-Kang et al., 2010), los modelos de ANN representan una compleja relación no lineal y extraen la dependencia entre las variables a través del proceso de formación. Los métodos basados en ANN incluyen redes neuronales de retropropagación, redes neuronales recurrentes, redes neuronales de función radial (RBF), red neuronal ridgelet, red neural de elementos lineales adaptativos, entre otros.

La Figura 2.3 muestra un ejemplo de la arquitectura de una red neuronal artificial, la cual esta compuesta por una capa de entrada, una capa oculta (puede tener una o más capas ocultas), y una capa de salida que puede estar definida por una o más salidas.

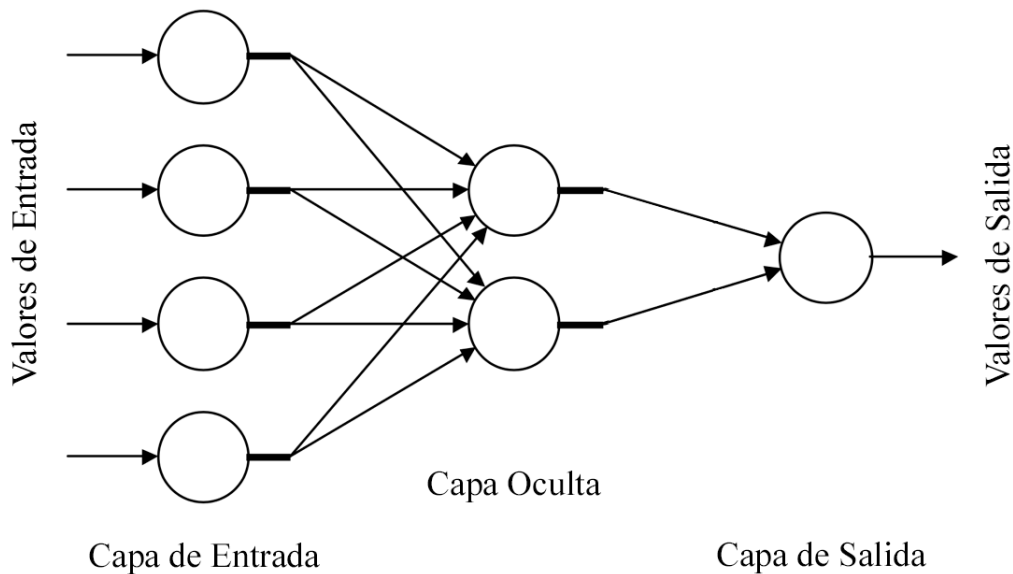


Figura 2.3: Imagen que representa un ejemplo de la arquitectura de una red neuronal artificial.

2.3.2.2. k - vecinos cercanos

El método de aproximación k vecinos cercanos (k -NN, por sus siglas en inglés) es muy simple pero poderoso; se ha utilizado en muchas aplicaciones, particularmente en tareas de clasificación y predicción (Sorjamaa and Lendasse, 2006). La idea de k -NN no es determinar el número de vecinos deseados, sino recuperar todos los que estén en su vecindad (aproximación), a una distancia de la observación actual.

Para resolver el problema de predicción, el método k -NN usa el promedio de las predicciones de los vecinos de la observación actual, sin hacer ninguna suposición acerca de la distribución de las variables a predecir en el proceso de aprendizaje.

La idea detrás del pronóstico k -NN es que ejemplos de entrenamiento similares muy probablemente tendrán futuros similares. Se debe buscar a todos los vecinos cercanos a la observación actual, de acuerdo a un criterio de distancia. La distancia que comúnmente se usa es la distancia Euclideana. Sin embargo, el método también permite el uso de otras distancias, como: Chebyshev, Manhattan y Mahalanobis, entre otras (Baldi and Brunak, 2001).

Una vez que se encuentra los vecinos más cercanos, se calcula una estimación de la variable de salida (el pronóstico) utilizando alguna técnica de fusión (i.e. el promedio, la mediana, entre otras) de las salidas de los vecinos más cercanos.

Este algoritmo es una técnica de clasificación y regresión basada en la similaridad de los individuos de una población. La premisa de esta técnica indica que los miembros de una población coexisten rodeados de individuos similares que tienen propiedades similares. Esta idea sencilla es la regla de aprendizaje del los clasificador y regresores k -NN. Como se muestra en la Figura 2.4 donde se busca la similaridad de los individuos más cercanos.

2.3.2.3. Regresión lineal

El análisis de regresión es una de las técnicas más utilizadas para analizar datos multifactoriales. Su utilidad resulta del proceso lógico de una ecuación para expresar la relación entre una variable de interés y un conjunto de variables predictoras relacionadas. Además, el análisis de regresión es interesante debido a las matemáticas subyacentes elegantes y una teoría estadística bien desarrollada. El uso exitoso de la regresión requiere una apreciación

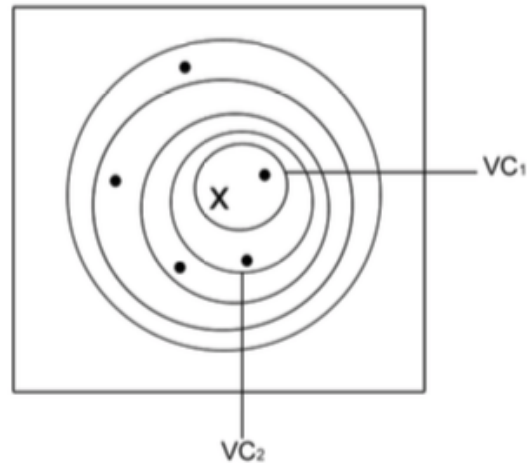


Figura 2.4: Ejemplo de clasificación basada en la similitud de los individuos k -NN

tanto de la teoría como de los problemas prácticos que típicamente surgen cuando la técnica se emplea con datos del mundo real (Montgomery et al., 2012).

La aplicación de regresión lineal se utiliza en diferentes puntos de vista práctico. A pesar de que el modelo se ha planteado para analizar la relación entre variables cuantitativas, su uso en regresión cualitativa es inmediata (Pértega Díaz and Pita Fernández, 2000).

El modelo de regresión lineal es sencillo para analizar la relación entre dos variables cuantitativas. Pero en la mayoría de los casos predice una respuesta en función de un conjunto de variables mas grande. En particular, el análisis de regresión consiste en construir modelos matemáticos que describen las relaciones que pueden existir entre las variables (Seber and Lee, 2012). La Figura 2.5 muestra un ejemplo de predicción aplicando regresión lineal.

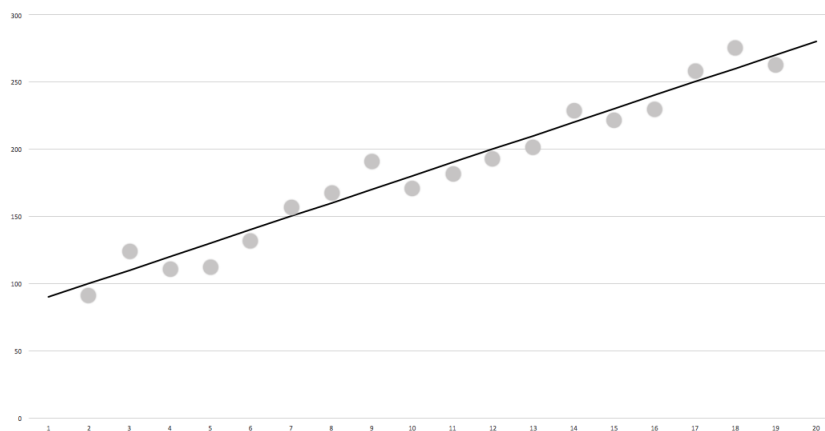


Figura 2.5: Imagen representativa del pronóstico aplicando regresión lineal.

2.3.2.4. Árboles aleatorios

EL modelo de árboles aleatorios es una combinación de árboles predictores. Cada árbol depende de valores de un vector aleatorio que se muestra de forma independiente pero con la misma distribución para todos los demás árboles utilizados (Breiman, 2001). La Figura 2.6 muestra la implementación de árboles aleatorios.

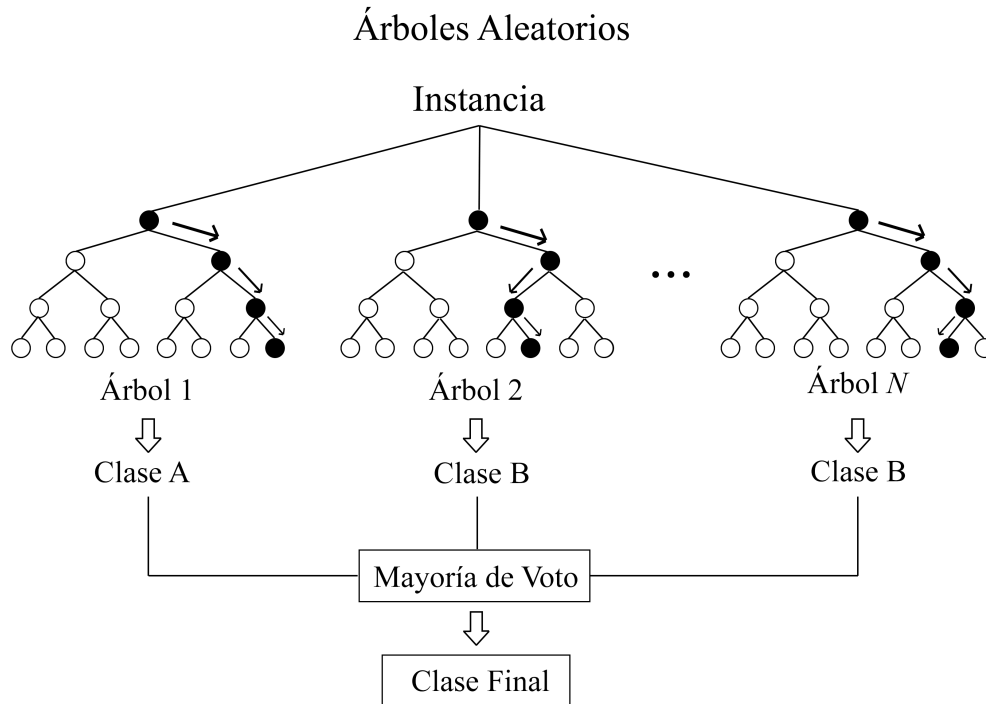


Figura 2.6: Imagen representativa de la implementación de árboles aleatorios en pronósticos.

La técnica de árboles aleatorios consiste en una colección de clasificadores estructurados por árboles. Además de construir cada árbol con diferentes muestra de datos, cambian la forma en que se construyen. En un árboles aleatorios, cada nodo se divide utilizando el mejor de un subconjunto de predictores elegidos al azar en ese nodo. Esta estrategia resulta funcionar muy bien en comparación con muchos otros clasificadores (Liaw et al., 2002).

Además, es muy fácil de usar en el sentido de que tiene solo dos parámetros (el número de variables en el subconjunto aleatorio en cada nodo y el número de árboles en el bosque), y por lo general no es muy sensible a sus valores. El pronosticador de árboles aleatorio utilizado para este estudio consiste en usar características seleccionadas al azar o una combinación de

características en cada nodo para hacer crecer un árbol (Pal, 2005).

El diseño de un árbol de decisión requirió la elección de una medida de selección de atributo y un método de poda. Hay muchos enfoques para la selección de atributos utilizados para la inducción del árbol de decisión y la mayoría de los enfoques asignan una medida de calidad directamente al atributo. El pronosticador de árboles aleatorios usa el índice de Gini como una medida de selección de atributo, que mide la impureza de un atributo con respecto a las clases. Este pronosticador implica elegir un conjunto de características al azar y crear un pronosticador con una muestra inicial de los datos de entrenamiento. Se genera una gran cantidad de árboles y, se hace una votación no ponderada para asignar un valor desconocido a una clase (Pal, 2005).

2.3.3. Métodos híbridos

El objetivo de los modelos híbridos es beneficiarse de las ventajas de cada modelo y obtener un rendimiento de predicción globalmente óptimo (Wu and Hong, 2007). Dado que la información contenida en el método de predicción individual es limitada. Los métodos híbridos maximizan la información disponible, integran la información del modelo individual y aprovechan al máximo las ventajas de los métodos de predicción múltiples, mejorando así la precisión de la predicción. Los métodos híbridos combinan diferentes enfoques, como mezclar enfoques físicos y estadísticos o combinar modelos a corto y mediano plazo (Soman et al., 2010).

2.4. Fusión de Datos

La teoría de decisiones o fusión de datos (FD) y la toma de decisiones adecuadas juegan un papel muy importante en el área de la fusión de decisiones, principalmente debido a las situaciones complejas encontradas en los procesos de fusión de datos multi-sensoriales (MS).

FD trata principalmente sobre cómo tomar decisiones y qué decisiones tomar sobre el estado de un objeto, escenario y/o evento, en base a algún análisis objetivo (a veces análisis subjetivo) llevado a cabo a partir de los conjuntos de datos. Las decisiones se toman en base a los datos disponibles de los sensores y muchas otras fuentes. A menudo estos conjuntos de

datos se procesan y luego se toman o toman algunas decisiones apropiadas.

El FD moderno cuenta con contribuciones de varias disciplinas, y existe una gran superposición, y el tema de FD se ha beneficiado de diversos métodos enfoques seguidos por investigadores de diferentes contextos (Raol, 2015). En este trabajo se utilizaron tres esquemas de fusión, la mediana aritmética, la media aritmética y un agrupamiento de datos, los cuales serán detallados a continuación.

2.4.1. Mediana Aritmética

Una de las técnicas que se utilizan en la fusión de datos es la mediana. La mediana es el valor de la variable estadística que divide en dos efectivos iguales a los individuos de un conjunto de datos ordenados por valor creciente del carácter (Cobo and Batanero, 2000).

Si n es impar, la mediana es el valor que ocupa la posición $\frac{n+1}{2}$ una vez que los datos han sido ordenados, porque éste es el valor central. La Ecuación 2.1 muestra la fórmula del cálculo de la mediana cuando n es impar.

$$m_e = x_{\frac{n+1}{2}} \quad (2.1)$$

Si n es par, la mediana es la media aritmética (promedio) de los dos valores centrales. Cuando n es par, los dos datos que están en el centro de la muestra ocupan las posiciones $\frac{n}{2}$ y $\frac{n}{2} + 1$. La Ecuación 2.2 muestra la fórmula del cálculo de la mediana cuando n es par.

$$m_e = \frac{(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})}{2} \quad (2.2)$$

2.4.2. Media Aritmética (Promedio)

Además de la mediana, en la fusión de datos se implementa la media aritmética o promedio. La media se define como la suma ponderada de cada uno de los valores de la variable, multiplicado por su frecuencia. Esta otra definición de la media, es el promedio aritmético de un conjunto de datos relaciona la media con otros promedios y enfatiza su carácter de valor central (Cobo Merino and Batanero, 2004). La Ecuación 2.3 muestra la fórmula del cálculo de la media aritmética.

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (2.3)$$

donde \bar{x} es el valor resultante del calculo del promedio de un vector, x_n es el valor correspondiente a un vector y n es el tamaño del vector.

2.4.3. Agrupamiento de Datos

Otra implementación utilizada en la fusión de datos, es una técnica de agrupamiento espacial. Esta técnica (en inglés *Density-based spatial clustering of applications with noise*), es un algoritmo de agrupamiento de datos propuesto por Martin Ester, Hans-Peter Kriegel, Jörg Sander y Xiaowei Xu en 1996 (Ester et al., 1996).

Es un algoritmo de agrupamiento basado en densidad porque encuentra un número de grupos ó *clusters* comenzando por una estimación de la distribución de densidad de los nodos correspondientes. Además, es uno de los algoritmos de agrupamiento más usados y citados en la literatura científica.

En este trabajo, el algoritmo de agrupamiento de datos se define una ecuación que da como resultado el valor pronosticado de una serie de tiempo, el objetivo de esta técnica es obtener los valores mas cercanos al valor mediana de un conjunto de datos. La Ecuación 2.4 muestra la fórmula definida para el pronóstico de series de tiempo en este trabajo.

$$y_{t+1} = Promedio(|M \pm C * i|) \quad (2.4)$$

donde $|M \pm C * i| > 0$, M es la mediana aritmética y $C * i$ es el *radio* permitido a considerar en el pronóstico.

Como se podrá ver, en la Figura 2.7 se puede ilustrar el rango permitido a considerar para la realización de la media aritmética, que será el pronóstico de las series de tiempo (y_{t+1}).

En la Figura 2.8 se muestra un ejemplo de como queda definido una agrupación de datos, donde los puntos negros representan los valores mas agrupados de un determinado conjunto de datos.

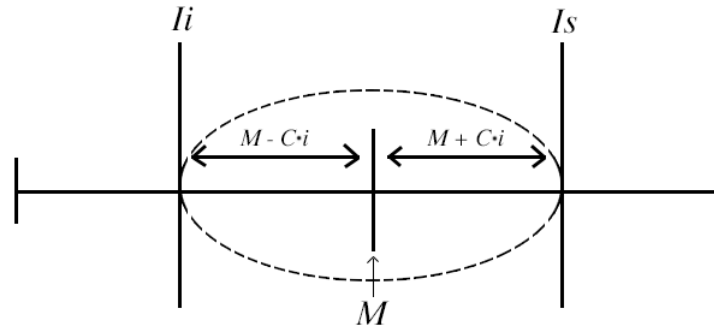


Figura 2.7: Ilustración del rango permitido para calcular el promedio de datos para el pronóstico de series de tiempo.

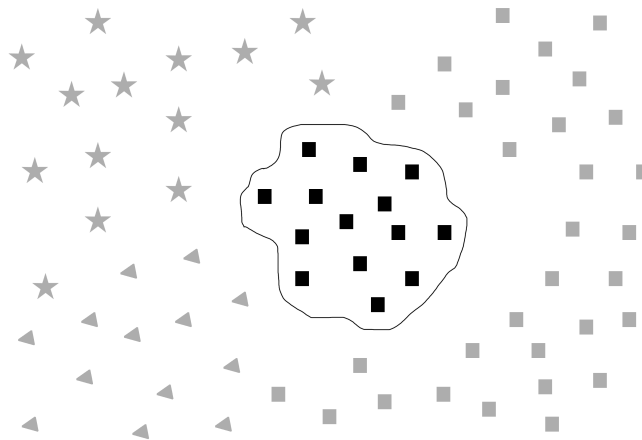


Figura 2.8: Muestra de elementos agrupados de un conjunto de datos.

Capítulo 3

Estado del arte

En esta sección se describen algunos trabajos de investigación importantes relacionados en el campo de series de tiempo y de los métodos de predicción. Estos trabajos guardan similitud con el presente trabajo de investigación y fueron considerados como punto de partida para el desarrollo de esta propuesta de investigación.

3.1. Pronóstico

Hoy en día la tarea de pronosticar ha generado grandes beneficios en diferentes áreas de la investigación. En el caso particular de la predicción de series de tiempo de la velocidad del viento. El beneficio generado ha sido mejorar la producción de energía, esto es, reducción de costos de producción y planeación de energía tomada de diferentes estaciones meteorológicas, teniendo como resultado una reducción del daño hacia el medio ambiente.

En la literatura se ha mencionado que hay una variedad de enfoques para realizar pronóstico de series de tiempo de la velocidad del viento. En el trabajo de (Chang, 2014), menciona diferentes técnicas de predicción de series de tiempo que son utilizadas para diferentes períodos de tiempo; Chang define cuatro categorías para realizar predicciones en un período de tiempo (muy corto, corto, mediano y largo plazo).

En el trabajo de (Wu and Hong, 2007) se presentan cinco métodos para llevar a cabo los pronósticos de serie de tiempo de la velocidad del viento: el método de persistencia, el método físico, el método estadístico, el método de correlación espacial y los métodos de inteligencia artificial. Wu y Hong aplicaron el método de persistencia, el cual utiliza una suposición sim-

ple de que la velocidad del viento en un momento futuro será la misma velocidad que cuando se hace la predicción. Además, este método suele ser más exacto que otros métodos de predicción de la velocidad del viento en el caso de períodos a muy corto plazo, pero la precisión de este método degrada rápidamente cuando la escala de tiempo del pronóstico aumenta.

Otro de los métodos aplicados por Wu y Hong es el método físico, este se basa en la predicción del tiempo atmosférico utilizando datos meteorológicos como temperatura, presión, rugosidad de la superficie y obstáculos. Además, los métodos físicos se representan en los supercomputadores, ya que necesitan muchos cálculos matemáticos. También se encuentran los métodos estadísticos; estos métodos son fáciles de modelar y más baratos de desarrollar en comparación con otros métodos. Básicamente, el método estadístico es bueno para períodos de tiempo cortos, pero la desventaja de este método es que el error de predicción aumenta a medida que aumenta el tiempo de predicción.

Otros métodos encontrados en la literatura son los de correlación espacial, según (Lei et al., 2009), estos tienen en cuenta la relación espacial de la velocidad del viento de diferentes sitios. Los métodos de correlación espacial emplean las series de tiempo de la velocidad del viento desde un punto predicho y sus puntos vecinos para predecir la velocidad del viento. Finalmente se encuentran los métodos de inteligencia artificial (IA) los cuales han tenido mejores resultados en comparación con las demás técnicas previamente descritas. A continuación se presentan algunos trabajos de predicción realizados por diferentes autores aplicando los métodos de IA.

3.2. Métodos de predicción con IA

La IA ha aportado métodos que han tenido buenos resultados en la tarea de predicción. En la investigación desarrollada por (Wang et al., 2009) se realizó un enfoque relacionado con datos hidrológicos, en este trabajo se hace una predicción de datos hidrológicos implementando técnicas de IA y otras técnicas de predicción para hacer una comparación del resultado de las predicciones. Se aplicó modelos autorregresivos de media móvil, enfoques de redes neuronales artificiales, técnicas de sistema de inferencia difusa basadas en redes neuronales, modelos de programación genética y máquinas de soporte vectorial utilizando observaciones

a largo plazo de las descargas mensuales del flujo del río.

Wang utilizó cuatro medidas estándar cuantitativas de evaluación del rendimiento estadístico: el coeficiente de correlación (R), el coeficiente de eficacia Nash-Sutcliffe (E), la raíz cuadrada del error medio (RMSE) y el error porcentual absoluto medio (MAPE), para evaluar el rendimiento de varios modelos desarrollados. Los resultados indicaron que los métodos de IA obtuvieron el mejor rendimiento en la evaluación durante las fases de capacitación y validación del trabajo. A diferencia del trabajo de Wang, con respecto al trabajo de tesis, se utilizó la medida del error cuadrático medio (MSE) para evaluar la aptitud de los modelos de predicción.

Otras técnicas de IA que ha contribuido en la tarea de predicción es vecinos cercanos. En el artículo desarrollado por (Flores et al., 2015), habla sobre el pronóstico de la velocidad del viento utilizando métodos de IA: vecinos cercanos (NN, por sus siglas en inglés) y lógica difusa (LD). Se definió como *Fuzzy Nearest Neighbors* (FNN, por sus siglas en inglés); para ponderar de manera implícita la contribución de los diferentes vecinos al momento de realizar la predicción de las series de tiempo. Para lograrlo se hace mediante la función de membresía difusa de los términos lingüísticos (TL), que en este caso actuaron como una forma de distancia a la observación actual. Durante el proceso de entrenamiento se reúne todos los escenarios diferentes en las observaciones pasadas de las series de tiempo y se hace un mapeo para formar reglas difusas.

El modelo de reglas difusas determinado por FNN contiene un conjunto de reglas de inferencia difusas. Se realizaron los experimentos y se hizo una comparación de la aptitud del modelo FNN con respecto a los modelos auto-regresivo integrado de media móvil (ARIMA, por sus siglas en inglés) y NN. La medición de la aptitud de los modelos fue mediante las medidas de MAPE y MSE.

Flores et al. determinaron que FNN no gana en todos los casos, esta propuesta está en desarrollo y debe ser probado con series de tiempo que tengan datos más grandes, además, de realizar más experimentos con conjuntos difusos más grandes. Este método tiene como fortaleza reconocer el escenario presente como algo ocurrido en el pasado y utilizar las experiencias pasadas, expresadas en la serie de tiempo, para pronosticar lo que sucederá a corto plazo. Este trabajo sirvió como modelo para redefinir la técnica de k -NN que se desarrolló en

esta propuesta de investigación

En el trabajo de (Rodriguez et al., 2017), se plantean el tema de realizar predicciones a corto plazo con series de tiempo incompletas. Se ha dado el incremento en el interés hacia la transición global de la producción de energía limpia y por esta razón el poder pronosticar variables de energía es de suma importancia para etapas de planeación, administración y producción de energía. En este trabajo se enfoca en la velocidad del viento; dado que el viento es muy variable, presenta series de tiempo con comportamiento caótico. La consecuencia de este comportamiento es que al obtener los datos, se presentan datos falsos o incompletos, lo cual perjudica al momento de realizar el pronóstico. Se realizó el proceso de reconstrucción de las series de tiempo para después proceder al pronóstico de las series de tiempo ya reconstruidas.

Se exploró la reconstrucción de series de tiempo aplicando redes neuronales artificiales (ANN, por sus siglas en inglés), para posteriormente realizar el pronóstico a corto plazo de series de tiempo de la velocidad del viento implementando el método de k vecinos cercanos (k -NN). En esta metodología, se utilizaron series de tiempo recolectadas de diferentes localidades del estado de Michoacán y se hizo una comparación de la aptitud del pronóstico (MSE) utilizando ANN, k -NN y de persistencia (Naïve) en series de tiempo incompletas y reconstruidas.

Respecto a la técnica de ANN, en la investigación de (Sfetsos, 2002) se presenta un método de ANN enfocado a la predicción de la velocidad media del viento por hora usando el análisis de series de tiempo. La metodología propuesta tiene un beneficio adicional para la utilidad que tiene un nivel significativo de penetración eólica y utiliza intervalos en horas para los procedimientos operativos del sistema eléctrico, tales como el despacho económico y el compromiso de la unidad, este trabajo tiene relación con la implementación de ANN en la predicción de la velocidad del viento, ya que se utilizó la topología de *Multi-layer Perceptron*.

Dentro de los trabajos relacionados con los métodos de IA para realizar predicciones se encuentran árboles aleatorios (RF, por sus siglas en inglés). En el trabajo de (Kane et al., 2014) se realizó una comparación de resultados de predicciones utilizando RF y ARIMA. Se aplicaron estos métodos en datos de incidencia de brotes de influenza aviar altamente patógena (H5N1) en el país de Egipto, donde se demostró que RF supera al modelo ARIMA

en capacidad de realizar pronósticos. Se encontró que el modelo de RF es eficaz en predecir brotes de H5N1 en Egipto. El modelado RF para el pronóstico de series de tiempo proporciona una capacidad predictiva mejorada sobre los modelos de series de tiempo existentes para la predicción de brotes de enfermedades infecciosas.

Este resultado junto con los que muestran la concordancia entre los brotes de aves y humanos (Rabinowitz et al., 2012), proporciona un nuevo enfoque para predecir estos brotes peligrosos en las poblaciones de aves sobre la base de los datos existentes y de libre disponibilidad. Este análisis revela la estructura de serie de tiempo de la gravedad del brote de influenza aviar altamente patógena (H5N1) en Egipto. este trabajo fue de gran utilidad para esta propuesta de investigación, ya que sirvió como base en la implementación de la técnica de RF en el modelo de predicción.

3.3. Fusión de datos

Cada resultado de una predicción se ha comparado con otros métodos de pronóstico, con el objetivo de determinar si el método es bueno para un determinado problema. Se ha propuesto la implementación de fusionar diferentes métodos de predicción con la finalidad de mejorar el pronóstico de un caso en particular.

En el trabajo de (Ghosh et al., 2011) habla sobre como la estrategia de fusionar datos es una colaboración efectiva, en este caso, entre métodos de diagnósticos de fallas heterogéneas. Se estudió los beneficios de la estrategia de fusión de datos basados en la utilidad y evidencia al tomar decisiones. Las estrategias de fusión se basaron en una votación ponderada, bayesiana y de (Shafer, 1992), teniendo un mejor resultado a diferencia de un método simple en el caso de fallas en procesos químicos.

Otro trabajo acerca de fusión de predicción de datos es de (Haque et al.,2012), quienes presentaron un algoritmo híbrido basado en la transformación de onda (WT, por sus siglas en inglés) y una red difusa teórica predictiva de resonancia adaptativa (FA, por sus siglas en inglés). Este modelo se implementó para pronosticar la producción de la potencia de un parque eólico, tomando como base la información obtenida de la velocidad, dirección y temperatura del viento. Se hizo una comparación de este modelo con métodos de persistencia

y algunos modelos de computación suave (SCM, por sus siglas en inglés). Los resultados obtenidos por (WT + FA) fueron positivos, esto fue mediante la comparación de resultados tomando como medidas de aptitud el error porcentual absoluto medio (MAPE) y el error de la raíz media cuadrada (NRMSE). La estrategia del modelo híbrido para el pronóstico de energía eólica se aplicó en parques eólicos en la ciudad de Nuevo Brunswick, Canadá.

En la propuesta de (Cao and Cao, 2006) presentaron una red de retropropagación recurrente de onda (RBPN, por sus siglas en inglés) para pronosticar la radiación solar. Es una fusión de redes neuronales y de la transformación de onda. La RBPN es una red neuronal dinámica que tiene la opción de retro-alimentar los resultados totales o parciales. La red neuronal artificial es una herramienta poderosa en el pronóstico de la radiación solar. Con el fin de obtener una mayor precisión en el pronóstico, la red neuronal artificial y el análisis de ondas se combinaron para desarrollar un nuevo método de predicción de la radiación solar.

Los resultados de RBPN mostraron que para pronósticos de un día de anticipación, el RBPN fue mejor en comparación con la aplicación de la transformación de onda por si sola. Se utilizó la medida de aptitud del error medio cuadrático del pronóstico teniendo un 7.83 % de avance en la predicción aplicando el RBPN. Este método para el pronóstico de la radiación solar se aplicó con datos utilizados de estaciones meteorológicas de Baoshan, Shanghai.

Estos trabajos fueron de gran utilidad para esta propuesta de investigación, ya que sirvieron como base en la implementación de la fusión de datos con diferentes pronosticadores utilizados en el modelo de predicción.

Capítulo 4

Desarrollo

En este capítulo se presenta la propuesta del modelo de diseño de predicción de series de tiempo de la velocidad del viento. El proceso de predicción de la velocidad del viento es una tarea complicada ya que existen diferentes métodos de realizar la tarea de predicción. Entre los más destacados se encuentran los métodos estadísticos, los métodos de persistencia, y la inteligencia artificial (Chang, 2014). La presente tesis explora la implementación de técnicas de inteligencia artificial (IA).

El desarrollo de este trabajo consistió en procesar las series de tiempo a una base de datos; esto ayudará en la tarea del pronóstico. Con la base de datos definida se realizan las predicciones necesarias utilizando los métodos de IA: redes neuronales, vecinos cercanos, regresión lineal y árboles aleatorios. Con el resultado de cada uno de los pronosticadores, se aplicó un proceso de fusión de datos; esta fusión contempla tres opciones diferentes: mediana, promedio y agrupamiento de datos.

La finalidad de aplicar el enfoque de fusión es determinar si este esquema obtiene mejores resultados con respecto a un método individual. En la Figura 4.1, se muestra el proceso del pronóstico de la velocidad del viento que se definió en este trabajo. A continuación se describe a detalle el desarrollo del pronóstico de la velocidad del viento.

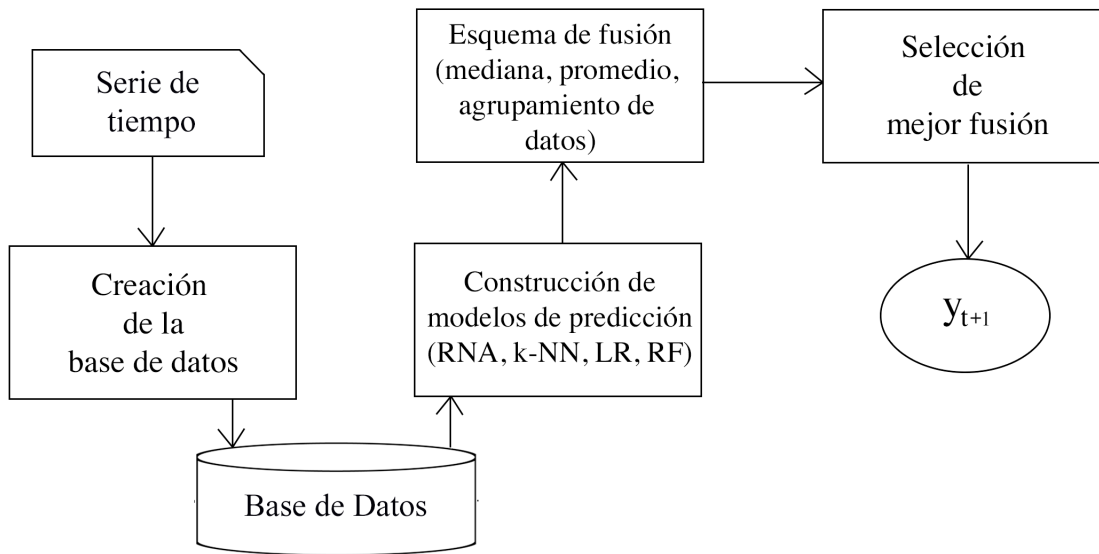


Figura 4.1: Proceso del pronóstico de la velocidad del viento.

4.1. Creación de la Base de Datos

Cada serie de tiempo presenta distintas características, tales como: distancia, tiempo, magnitud, etc. y modelar estas series de tiempo de la velocidad del viento requiere el ajuste de distintos parámetros, por lo que es necesario modelar cada una de ellas. Se determinó trasladar las series de tiempo a una base de datos.

El desarrollo de la base de datos para el pronóstico de series de tiempo consistió en seleccionar una muestra de cada serie de tiempo para el entrenamiento y otra para validación para obtener el pronóstico. En primera instancia, a la muestra seleccionada se le aplicó el proceso de normalización (ver en la subsección 4.1.1.1). A partir del proceso de normalización aplicada se definió un tamaño de ventana (m) asociada a la muestra de la serie de tiempo, para ser trasladada a la base de datos.

4.1.1. Procesamiento de los datos

Las series de tiempo de la velocidad del viento son un conjunto de datos históricos que se recolectan a través del tiempo. Se definió un tamaño de ventana (m) la cual se asocia a una muestra tomada de la serie de tiempo para desarrollar los modelos de predicción. La ventana

se definirá realizando diferentes iteraciones para determinar un tamaño óptimo; esta ventana de datos servirá en la tarea del pronóstico de la velocidad del viento. La Figura 4.2 representa el tamaño de ventana obtenida de cada serie de tiempo.

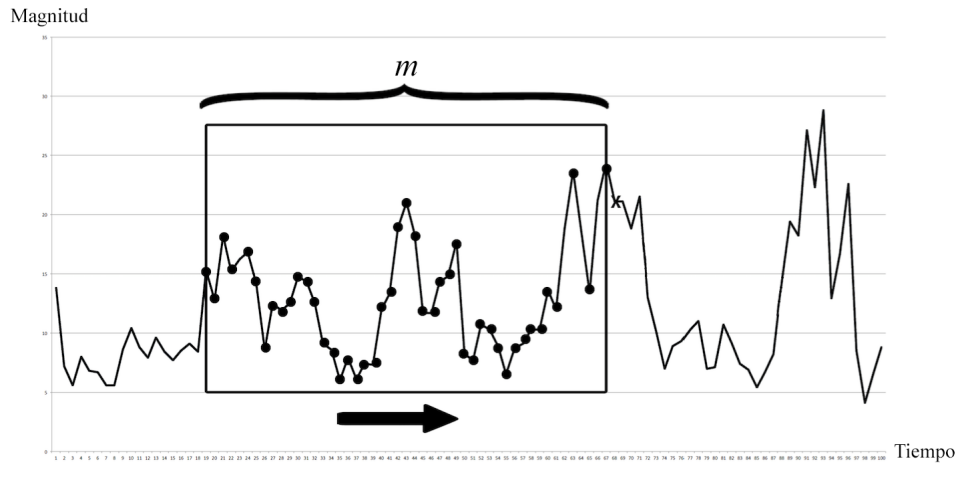


Figura 4.2: Muestra del tamaño de ventana de una serie de tiempo.

Una vez definida el tamaño de ventana se procesan los datos a una base de datos. Antes de esta tarea, se debe realizar un proceso de normalización de datos para continuar con la predicción de la velocidad del viento, este proceso de normalización se describe a continuación.

4.1.1.1. Normalización

El proceso de normalización consiste en designar y aplicar una serie de reglas a las relaciones obtenidas tras el paso del modelo entidad-relación al modelo relacional. En este caso, se aplicó una serie de reglas a la muestra seleccionada de cada serie de tiempo para después distribuir los datos normalizados a la base de datos.

Además, el objetivo de realizar la normalización de la serie de tiempo es tener los valores de los intervalos de hora de cada serie de tiempo a un mismo nivel, es decir, ser medibles para el proceso de entrenamiento y validación en el rango de 0 y 1. La ecuación 4.1 especifica la fórmula de normalización de datos que se realizó en este trabajo.

$$v_n = \frac{v_r - v_{min}}{v_{max} - v_{min}} \quad (4.1)$$

donde v_n es el valor normalizado, v_r el valor real del intervalo de hora de la serie de tiempo,

v_{max} es el valor máximo de la muestra de datos de la serie de tiempo; y v_{min} el valor mínimo de la muestra de los datos de la serie de tiempo a pronosticar.

Con los datos normalizados de cada serie de tiempo, se aplica la tarea de pronosticar las serie de tiempo de la velocidad del viento. La Figura 4.3 muestra como los datos *crudos* de la serie de tiempo son normalizados para obtener valores entre 0 y 1.

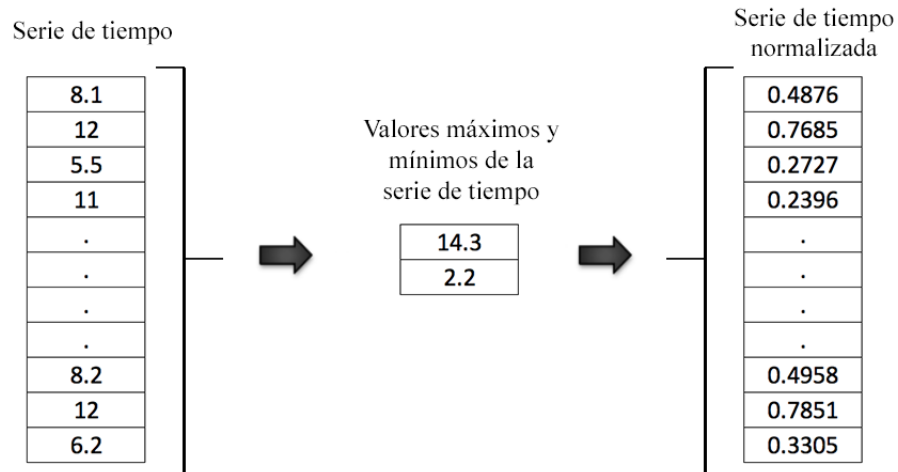


Figura 4.3: Proceso de normalización de las series de tiempo.

Al final, los datos normalizados se utilizarán en la tarea de predicción de series de tiempo de la velocidad del viento. Una vez obtenido los datos normalizados, se pasa a la tarea de procesar los datos a una base de datos para construir el modelo de predicción de series de tiempo de la velocidad del viento.

4.1.2. Generación de la base de datos

Los datos procesados (normalizados) se almacenan y distribuyen en una base de datos; esta base de datos registra las m (tamaño de ventana) características y las asocia con su valor respectivo (valor a pronosticar). La Tabla 4.1 muestra como se procesa la serie de tiempo a una base de datos y extrae las a características asociadas con un resultado esperado. Cada registro dentro de la base de datos representa las últimas m observaciones en un tiempo definido (t) asociado a una salida esperada $P(m) = a_{t+1}$ para el entrenamiento.

Tabla 4.1: Base de datos de la serie de tiempo.

	Entradas					Salida
						$P(m) = a_{t+1}$
1	$a_1,$	$a_2,$	$a_3,$	$a_4,$	\dots	a_m
2	$a_2,$	$a_3,$	$a_4,$	$a_5,$	\dots	a_{m+1}
3	$a_3,$	$a_4,$	$a_5,$	$a_6,$	\dots	a_{m+2}
4	$a_4,$	$a_5,$	$a_6,$	$a_7,$	\dots	a_{m+3}
5	$a_5,$	$a_6,$	$a_7,$	$a_8,$	\dots	a_{m+4}
6	$a_6,$	$a_7,$	$a_8,$	$a_9,$	\dots	a_{m+5}
7	$a_7,$	$a_8,$	$a_9,$	$a_{10},$	\dots	a_{m+6}
8	$a_8,$	$a_9,$	$a_{10},$	$a_{11},$	\dots	a_{m+7}
9	$a_9,$	$a_{10},$	$a_{11},$	$a_{12},$	\dots	a_{m+8}
10	$a_{10},$	$a_{11},$	$a_{12},$	$a_{13},$	\dots	a_{m+9}
			\dots			
n	$a_{n-m-1},$	$a_{n-m},$	$a_{n-m+1},$	$a_{n-m+2},$	\dots	a_{m-1}

A partir de la creación base de datos, que almacena los datos normalizados, es posible construir los modelos de predicción de la serie de tiempo; para esto se requiere implementar las técnicas de inteligencia artificial seleccionadas. Además se lleva a cabo una fusión de datos utilizando los resultados de los pronosticadores mediante medidas estadísticas, tales como: mediana, promedio y agrupamiento de datos.

4.2. Construcción de modelos de predicción

Un paso importante en el desarrollo del trabajo de predicción es la construcción del modelo de predicción, ya que a partir de él es posible realizar un pronóstico de la serie de tiempo, este proceso se ilustra en la Figura 4.4. Para la elección del modelo que reconstruya la serie de tiempo del viento se compararon diferentes algoritmos de inteligencia artificial (IA), tales como: redes neuronales artificiales (ANN), el algoritmo de k vecinos más cercanos (k-NN), regresión lineal (LR) y árboles aleatorios (RF).

Cada método de IA fue implementado de manera individual con la finalidad de realizar una fusión de datos con los valores pronosticados. La Figura 4.4 muestra el proceso general utilizado por cada uno de los métodos de predicción de IA empleados para realizar el pronóstico de la velocidad del viento.

El proceso de fusión de datos consistió en implementar un esquema de fusión con tres

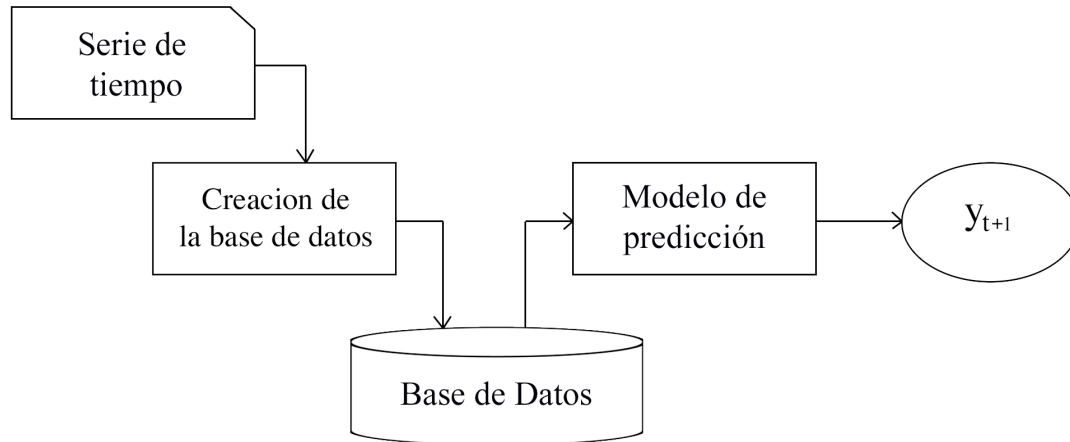


Figura 4.4: Proceso del pronóstico de la velocidad del viento definido para cada técnica de predicción.

diferentes opciones: 1) cálculo de la mediana aritmética, 2) media aritmética (promedio) y 3) el agrupamiento de datos entre los pronosticadores. Con base a esto, se obtendrán nuevas predicciones para ser comparados con los resultados individuales obtenidos de los modelos simples.

La Figura 4.5 describe el proceso del esquema de fusión. Con este proceso de fusión se obtendrá el valor pronosticado (y_{t+1}). Esto permitirá realizar una comparación de resultados entre los resultados individuales de cada método y los obtenidos por los diferentes métodos de fusión para elegir el más adecuado para realizar el pronóstico con base en el cálculo del error cuadrático medio (MSE, por sus siglas en inglés).

4.2.1. Pronóstico con aplicando redes neuronales

En el caso del pronóstico utilizando ANN, la arquitectura implementada en la red fue un *Feedforward Multilayer Perceptron* (Negnevitsky, 2005). El entrenamiento fue utilizando métodos basados en gradiente. La Figura 4.6 muestra la arquitectura general de la red neuronal implementada para el pronóstico.

La red neuronal esta compuesta por un conjunto de m observaciones pasadas, consideradas como los datos de entrada; la capa oculta consta de h neuronas y la capa de salida corresponde al valor pronosticado y_{t+1} ; se utilizó un *sigmoide* como función de activación.

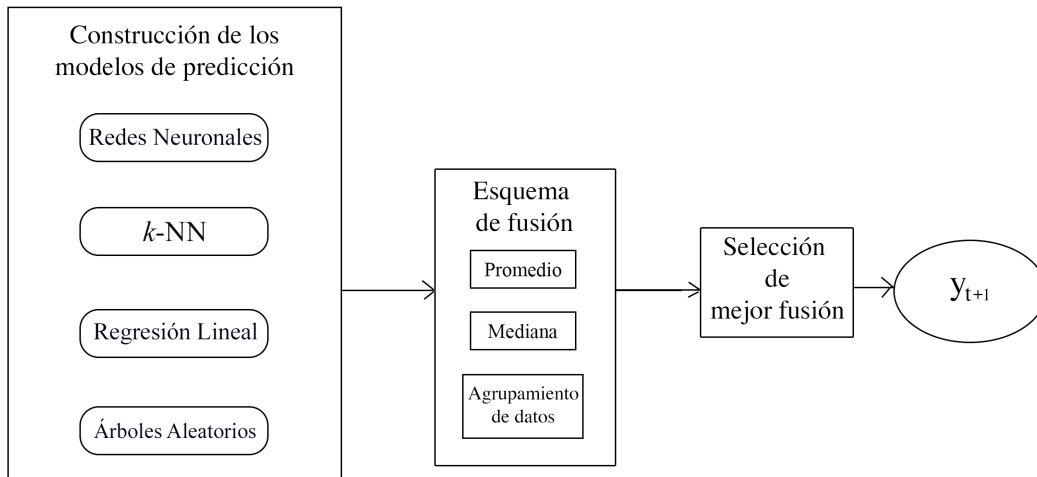


Figura 4.5: Implementación del esquema de fusión con los modelos de predicción utilizados en este trabajo.

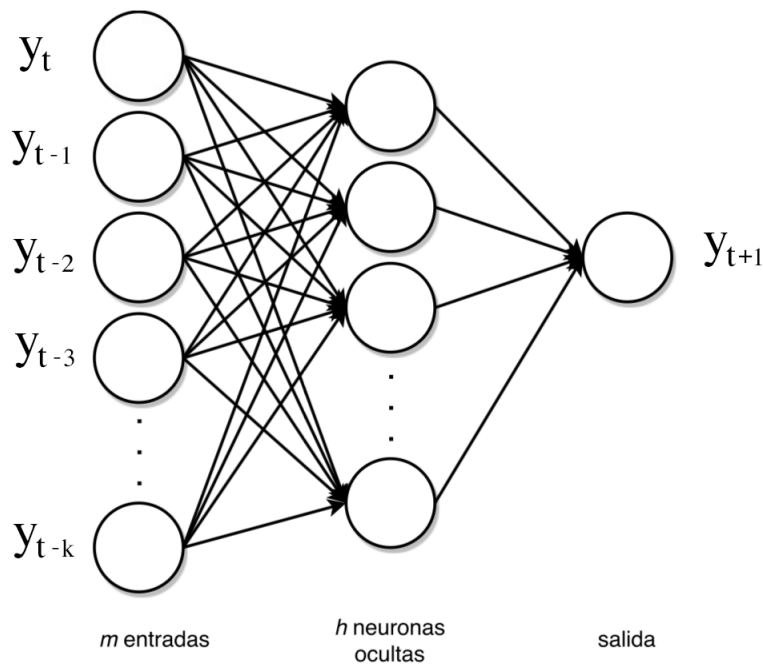


Figura 4.6: Arquitectura de ANN con m observaciones pasadas, h neuronas en su capa oculta y una única salida (y_{t+1}).

La red neuronal que modela la serie de tiempo de la velocidad del viento y_{t+1} esta definida por la Ecuación 4.2:

$$\begin{aligned}\hat{y}_{t+i} &= f_1\left(\sum_l^w w_l x_l\right) \\ x_l &= f_2\left(\sum_j^m w_{lj} y_{t-l_j}\right)\end{aligned}\tag{4.2}$$

donde f_1 y f_2 son las funciones de activación, y w son los coeficientes (también concidos pesos sinápticos).

Para encontrar la topología adecuada de la red neuronal se aplicó una iteración de valores en un rango de 2 a 41, ya que a partir del 41 el rendimiento no era adecuado. Este experimento determinó el número de neuronas de entrada m , número de neuronas h requeridas en la capa oculta que ejecutará en la red neuronal.

4.2.2. Pronóstico con k -vecinos cercanos

El algoritmo k -NN requiere buscar la similitud del individuo x en la base de datos definida. Donde el individuo x es definido como las observaciones pasadas de la serie de tiempo en un tiempo t ($x = [y_{t-w}, y_{t-w+1}, \dots, y_t]$). Después de evaluar las distancias de los individuos de la base de datos, se seleccionan los k vecinos más cercanos y el pronóstico resultante es el promedio de la columna y_{t+1} de la base de datos.

La definición matemática del modelo de k -NN se observa en la Ecuación 4.3.

$$y_{t+1} = \frac{1}{k} \sum_{i=1}^k P(VC(i, y_t))\tag{4.3}$$

La distancia euclidiana de dos individuos se define por la Ecuación 4.4.

$$d(r, s) = \sqrt{\sum_{j=1}^m (r_j - s_j)^2}\tag{4.4}$$

La ventana que termina en el tiempo t , denotada por w_t es la secuencia o vector de las m mediciones hasta t como se indica en la Ecuación 4.5.

$$w_t = [y_{t-m+1}, y_t] \in \mathbb{R}^m\tag{4.5}$$

La definición recursiva del conjunto de k vecinos más cercanos se observa en la Ecuación

4.6:

$$VC(k, x) = \operatorname{argmin}(d(w, x) | w \in W - V_{k-1}(x)) \quad (4.6)$$

donde $V_k(x)$ se especifica en la Ecuación 4.7:

$$V_k(x) = \bigcup_{j=0}^k VC_j(k, x) \quad (4.7)$$

$$V(0, x) = \phi$$

Para la determinación de los valores adecuados para k y m se realizó una exploración en un rango de valores lo suficientemente amplio para establecer las combinaciones que permitan minimizar el error. En este caso se realizaron experimentos en donde se iteraron estos valores desde $k = 2$ hasta $k = 41$ y desde $m = 2$ hasta $m = 41$.

4.2.3. Pronóstico con regresión lineal

La técnica de regresión lineal (LR, por sus siglas en inglés) predice la salida del modelo usando una combinación lineal de características (entradas) numéricas $x = [x_1, x_2, \dots, x_n]$ (Wolfram et al., 1996). La probabilidad condicional $P(y|x)$ está modelada de acuerdo a la Ecuación 4.8 (considerando la función 4.9 como parámetro):

$$P(y|x) \propto \exp(-(y - f(\theta, x))^2 / (2\sigma^2)) \quad (4.8)$$

$$f(\theta, x) = x \cdot \theta \quad (4.9)$$

La estimación del vector del parámetro θ se realiza minimizando la función de pérdida mostrada en la Ecuación 4.10:

$$\frac{1}{2} \sum_{i=1}^m (y_i - f(\theta, x_i))^2 + \lambda_1 \sum_{i=1}^n |\theta_i| + \frac{\lambda_2}{2} \sum_{i=1}^n \theta_i^2 \quad (4.10)$$

donde m es el tamaño de ventana de la serie de tiempo y n es el número de características numéricas.

Para este proceso se generaron datos bidimensionales considerados desde la ventana de la serie de tiempo. Además, se aplicó una función de entrenamiento para realizar la predicción de los datos. En la Figura 4.7 se muestra gráficamente como los datos bidimensionales están esparcidos. A partir, de los resultados mostrados en la Figura 4.8 se realizó una comparación entre los datos con los valores predichos observando su desviación estándar.

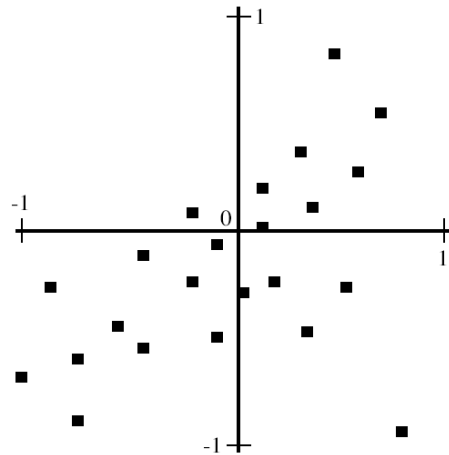


Figura 4.7: Datos bidimensionales esparcidos gráficamente.

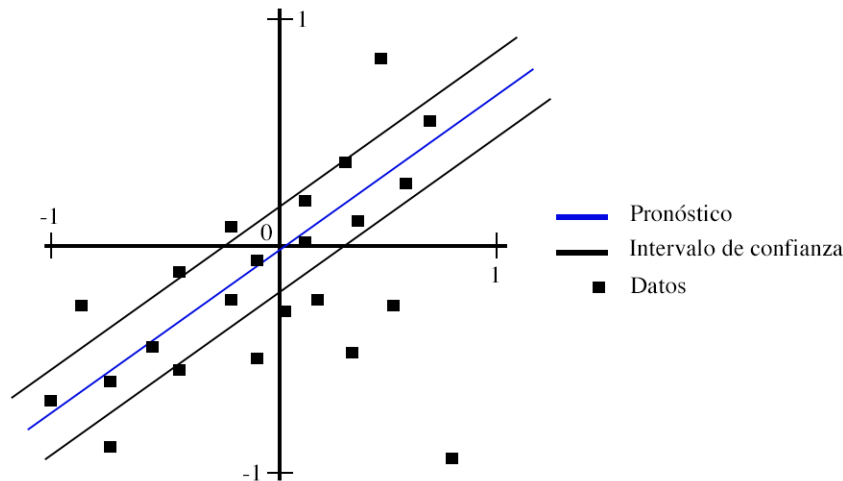


Figura 4.8: Predicción de la serie de tiempo aplicando regresión lineal.

En el proceso de predicción de las series de tiempo de la velocidad del viento con LR, consistió en realizar iteraciones en el entrenamiento para obtener el mejor tamaño de ventana m y el tamaño de n .

4.2.4. Pronóstico con árboles aleatorios

Los árboles aleatorios (RF) son un método de aprendizaje realizado por conjuntos para la clasificación y la regresión que opera mediante la construcción de una multitud de árboles de decisión. En este trabajo la predicción de RF se obtiene tomando las clases más comunes o las predicciones del árbol de valor medio; cada árbol de decisión se entrenó con un subconjunto aleatorio generado a partir del conjunto de entrenamiento y solo se utilizó un subconjunto aleatorio de las características (Wolfram et al., 1996).

El proceso aleatorio utilizado por el algoritmo de RF se manifiesta en la elección, tanto de la línea del recuadro (de observaciones) como de los indicadores. Precisamente, este proceso determina la resistencia significativa al ruido, irrupción y reentrenamiento, si lo comparamos con otros clasificadores en forma de árbol.

Con la Figura 4.9 se compararon los datos con los valores predichos observando su desviación estándar .

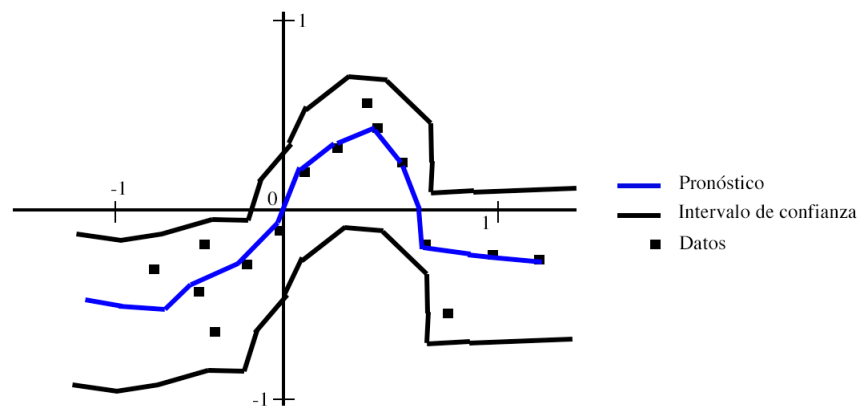


Figura 4.9: Proceso del pronóstico de series de tiempo de la velocidad del viento aplicando árboles aleatorios.

En el proceso de ajuste para la predicción de las series de tiempo de la velocidad del viento con RF consistió en realizar iteraciones durante el entrenamiento para obtener el tamaño adecuado de la ventana m y el tamaño de n .

4.3. Esquema de fusión

Con el resultado de los pronósticos obtenidos con las distintas técnicas de IA, se implementó un esquema de fusión para obtener un mejor pronóstico de precisión.

Este esquema contempla tres diferentes métodos para evaluar la unión de los resultados obtenidos por los métodos individuales de cada técnica de IA seleccionada e implementada en esta tesis para la predicción de la velocidad del viento. Las técnicas de fusión utilizadas fueron: 1) cálculo de la mediana, 2) el promedio y 3) la técnica de agrupamiento de datos. Estos métodos se describe en las siguientes secciones.

4.3.1. Implementación de la mediana

El proceso de fusión basado en la mediana para realizar el pronóstico fue realizado con las predicciones obtenidas de las técnicas: ANN, k -NN, LR, y RF. La ecuación 2.1 describe el proceso de obtención de la mediana que se aplicó en la fusión de los datos pronosticados. En este caso si el tamaño (n) de datos pronosticados es impar, se implementa dicha ecuación. Si n es par, la mediana es la media aritmética (promedio) de los dos valores centrales. Cuando n es par, los dos datos que están en el centro de la muestra ocupan las posiciones $\frac{n}{2}$ y $\frac{n}{2} + 1$. La Ecuación 2.2 muestra la fórmula del cálculo de la mediana cuando n es par.

En la Figura 4.10 se visualiza el proceso de la aplicación de la mediana con los diferentes pronosticadores.

En este esquema, se realizaron dos tipos de mediana, 1) utilizando el mejor resultado de la iteración de los métodos pronosticadores, el cual se denominó como Mediana Mejor, y 2) el calculo de todos los pronósticos obtenidos en la iteración, este se denominó como Mediana General.

4.3.2. Implementación de la promedio

El esquema de fusión por medio del promedio fue realizando con las técnicas de IA: ANN, k -NN, LR, y RF. Consistió en utilizar las predicciones que se obtuvieron en con dichas técnicas. La Ecuación 2.3 describe el proceso del promedio que se aplicó en la fusión de los

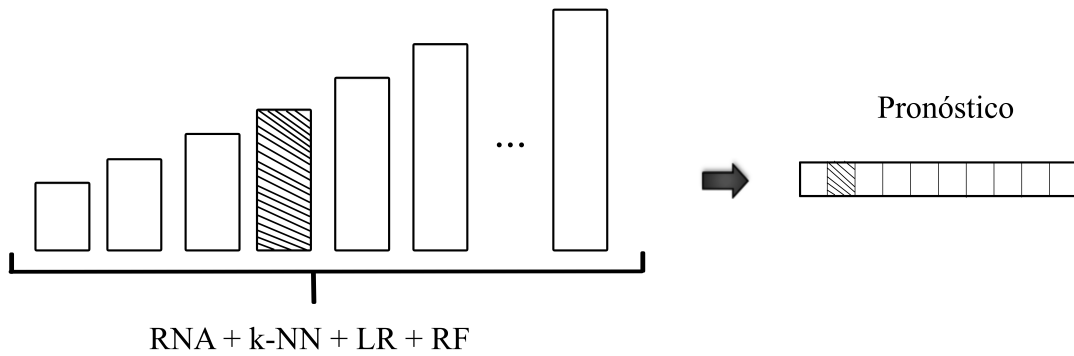


Figura 4.10: Proceso de la aplicación de la mediana con los diferentes pronosticadores.

datos pronosticados.

En la Figura 4.11 se visualiza el proceso de la aplicación del promedio con los diferentes pronosticadores.

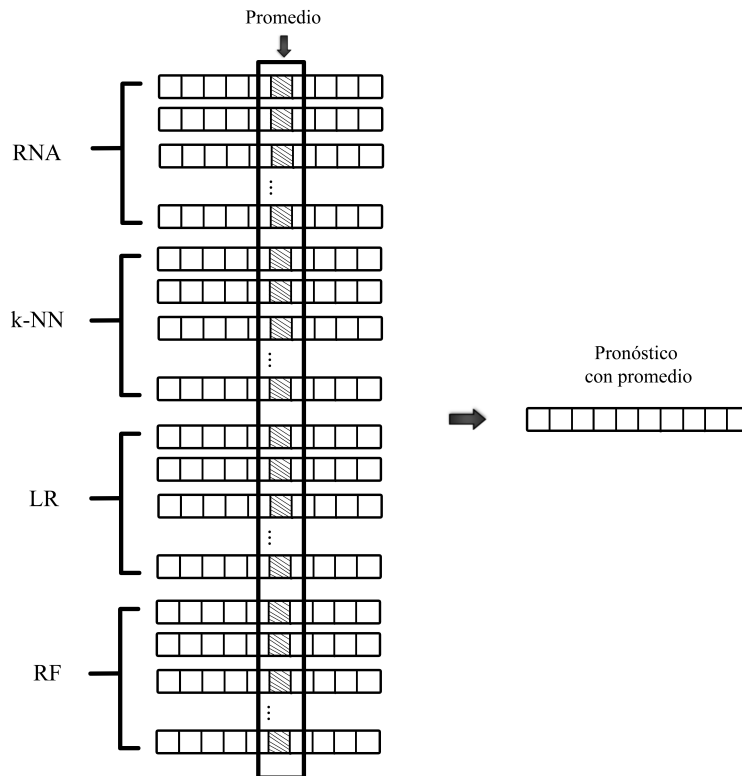


Figura 4.11: Proceso de la aplicación del promedio con los diferentes pronosticadores.

En este proceso de fusión, se realizaron tres implementaciones del promedio: 1) se calculó

con base en todos los pronósticos obtenidos en la iteración (ANN + k -NN + LR + RF), denominado Promedio General, 2) el promedio con los mejores resultados obtenidos en la iteración de cada pronosticador, denominado Promedio Mejor, y 3) una combinación del promedio con los pronósticos mejores con cada técnica.

4.3.3. Implementación de agrupamiento de datos

El método de agrupamiento de datos se definió considerando como base la mediana aritmética del resultado del pronóstico de series de tiempo. Durante el proceso de la mediana (V_m) del vector de la serie de tiempo, se obtuvo el valor cercano a la mediana, es decir, el valor siguiente de la serie de tiempo (V_{m+1}). La Figura 4.12 ejemplifica la forma en que es calculada.

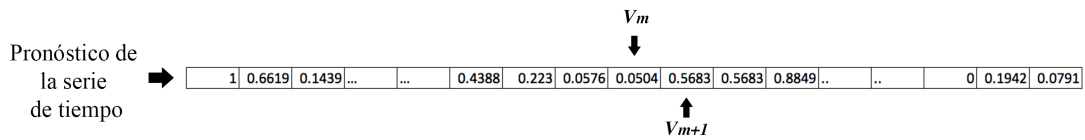


Figura 4.12: Valores de la serie de tiempo pronosticados con las técnicas de IA.

Estos dos valores fueron utilizados para obtener un nuevo vector con los valores más cercanos a la mediana. Para poder obtener el nuevo vector, se implementó la Ecuación 4.11.

$$\Delta = V_{m+1} - V_m \tag{4.11}$$

El valor (Δ) se utilizó para obtener el valor que definió los límites inferiores (I_i) y superiores (I_s) del nuevo vector. Se definió este valor como radio (R), esta fórmula se muestra en la Ecuación 4.12

$$R = C * \Delta \tag{4.12}$$

donde C se definió como un valor de aproximación para obtener el valor de R .

Con el valor de radio obtenido, se utilizó para obtener los valores de los límites inferiores y superiores del nuevo vector de los valores pronosticados.

La Ecuación 4.13 muestra como se obtuvieron los limites del nuevo vector.

$$I_i = V_m - R \tag{4.13}$$

$$I_s = V_m + R$$

Esto determino el tamaño del vector de datos con lo que se realizó un recorrido del vector actual para observar que datos del vector están dentro de los limites obtenidos con la Ecuación 4.13.

En la Figura 4.13 se muestra específicamente como se obtiene el nuevo vector de datos.

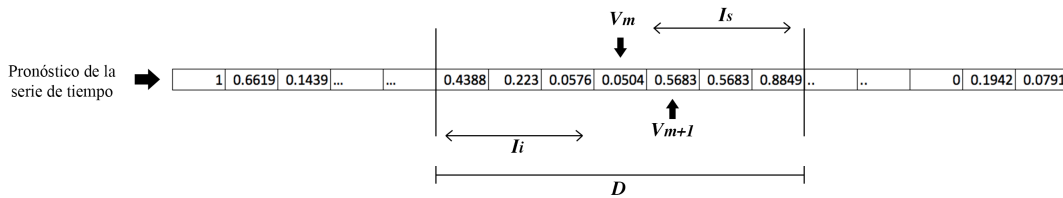


Figura 4.13: Proceso de cálculo de agrupamiento de datos para obtener el pronóstico de la serie de tiempo.

Donde D es el tamaño del vector datos de pronosticados para calcular el promedio y obtener el mse . I_i es el intervalo inferior, I_s es el intervalo superior, V_m es la mediana calculada, V_{m+1} es el valor que esta en la siguiente posición.

Con el nuevo vector, se calculó el promedio de ese vector para obtener el nuevo valor pronosticado. La Ecuación 4.14 muestra como se obtuvo la media aritmética.

$$y_{t+1} = \sum_{i=1}^n D_i \tag{4.14}$$

En este proceso de agrupamiento se realizó una iteracion de la variable C , para obtener el tamaño del vector de los datos agrupados con el fin de obtener el pronóstico de la serie de tiempo.

4.4. Selección de la mejor fusión

Con el resultado de las fusiones se realizó una comparación de los datos pronosticados con la mediana, promedio y del agrupamiento de datos, además de los métodos individuales.

Para poder seleccionar la fusión adecuada de los datos pronosticados se midió la aptitud de cada modelo y de cada fusión utilizando la medida de aptitud error cuadrático medio (MSE).

Capítulo 5

Análisis de los Resultados

En este capítulo se muestran los resultados obtenidos de los diferentes modelos propuestos para realizar el pronóstico de series de tiempo de la velocidad del viento recolectadas de diferentes localidades de Michoacán, México, incluyendo las propuestas para la fusión de resultados descritas en el capítulo 4. Además, se presenta una comparación entre los distintos resultados de los modelos y su representación gráfica para observar el comportamiento del pronóstico de la velocidad del viento.

5.1. Resultados

Se evaluó la aptitud de la propuesta metodológica de la presente tesis realizando diversos experimentos con series de tiempo de la velocidad del viento obtenidas de diferentes localidades eólicas dentro del estado de Michoacán, México. La forma de como se recolectaron los datos de las series de tiempo fue por medio de herramientas llamadas anemómetros.

Se realizaron predicciones de series de tiempo de la velocidad del viento con comportamiento caótico. En la tabla 5.1 se presenta las características caóticas de cada serie de tiempo de la velocidad del viento; donde λ_m es la medida que se encarga de medir que tan rápido se separan los datos uno sobre otro.

Un valor positivo de λ_m indica que el sistema dinámico que produjo las series de tiempo es caótico, por lo tanto, todas las series de tiempo utilizadas en este trabajo son caóticas. Estas series de tiempo tienen registros a intervalos horarios, es decir, cada dato es una hora en el tiempo (Montgomery et al., 2015).

Tabla 5.1: Series de tiempo con características caóticas

Serie de tiempo	λ_m
El Fresno	0.0720
La Palma	0.0759
Cointzio	0.0655
Mal Pais	0.0290
Corrales	0.0523

Se aplicó el proceso descrito en el capítulo 4: procesamiento de datos, normalización y creación la base de datos, para obtener las series de tiempo completas sin datos atípicos. Estos experimentos se desarrollaron utilizando la herramienta *Mathematica* en conjunto con el lenguaje de programación *Python* para generar el esquema de fusión de datos. En la parte de la fusión de datos utilizando promedio, se realizaron diferentes combinatorias con cada técnica de predicción. En la Tabla 5.2 se muestra la combinación del promedio con los métodos de IA (ANN, k -NN, LR y RF).

Tabla 5.2: Promedio con los diferentes métodos de IA

Promedio
ANN + k -NN + LR + RF
ANN + k -NN + LR
ANN + k -NN + RF
ANN + LR + RF
ANN + RF
ANN + LR
ANN + k -NN
k -NN + LR + RF
k -NN + LR
k -NN + RF
LR + RF

Los resultados obtenidos por los diferentes métodos de IA y el esquema de fusión con estos pronosticadores se muestran en la Tabla 5.3, donde se hace una comparación de errores definidos por la medida estadística MSE. Fueron en total siete series de tiempo del estado de Michoacán utilizadas en experimentos: Aristeo Mercado, El Fresno, La Palma, La Piedad, Cointzio, Malpais, y Patzcuaro.

Como puede mostrarse en la Tabla 5.3, el esquema de fusión obtuvo un resultado mejor

5. ANÁLISIS DE LOS RESULTADOS

Tabla 5.3: Resultados de los pronósticos en cada serie de tiempo

Model	Series de tiempo						
Resultados	A.Mercado	El Fresno	La Palma	La Piedad	Cointzio	Malpais	Patzcuaro
ANN	0.0048	0.0097	0.0095	0.0064	0.0055	0.0060	0.0004
<i>k</i> -NN	0.0056	0.0114	0.0097	0.0074	0.0061	0.0073	0.0006
LR	0.0050	0.0099	0.0098	0.0066	0.0059	0.0069	0.0005
RF	0.0049	0.0100	0.0094	0.0068	0.0057	0.0058	0.0004
Agrupamiento de datos	0.0050	0.0101	0.0095	0.0068	0.0057	0.0064	0.0005
Mediana							
General (ANN + <i>k</i> -NN + LR + RF)	0.0050	0.0101	0.0095	0.0068	0.0057	0.0064	0.0005
Mejor (ANN + <i>k</i> -NN + LR + RF)	0.0048	0.0094	0.0092	0.0062	0.0053	0.0059	0.0004
Promedio							
General (ANN + <i>k</i> -NN + LR + RF)	0.0050	0.0102	0.0095	0.0069	0.0058	0.0066	0.0005
Mejor (ANN + <i>k</i> -NN + LR + RF)	0.0047	0.0096	0.0092	0.0062	0.0053	0.0059	0.0004
(ANN + <i>k</i> -NN + LR)	0.0048	0.0097	0.0094	0.0063	0.0054	0.0062	0.0004
(ANN + <i>k</i> -NN + RF)	0.0047	0.0096	0.0091	0.0062	0.0052	0.0058	0.0004
(ANN + LR + RF)	0.0047	0.0094	0.0093	0.0062	0.0053	0.0058	0.0004
(ANN + LR)	0.0048	0.0095	0.0095	0.0063	0.0054	0.0062	0.0004
(ANN + RF)	0.0046	0.0094	0.0091	0.0062	0.0053	0.0056	0.0004
(ANN + <i>k</i> -NN)	0.0049	0.0100	0.0094	0.0064	0.0053	0.0061	0.0004
(<i>k</i> -NN + LR + RF)	0.0049	0.0098	0.0093	0.0064	0.0054	0.0062	0.0005
(<i>k</i> -NN + LR)	0.0051	0.0102	0.0095	0.0066	0.0058	0.0067	0.0005
(<i>k</i> -NN + RF)	0.0050	0.0100	0.0092	0.0065	0.0054	0.0061	0.0005
(LR + RF)	0.0048	0.0096	0.0094	0.0064	0.0055	0.0060	0.0005

en comparación con un método simple de IA en todas las series de tiempo. Específicamente la opción de utilizar el promedio fue el que obtuvo mejor resultado.

Haciendo una comparación estadística, la diferencia de resultados entre el esquema de fusión y el resultado de métodos simples, se obtuvo un 0.025% de mejoramiento en la predicción. Tal vez la diferencia se ve muy pequeña, pero si se hace una comparación en términos de reducción de costos de producción de la energía eólica, la diferencia del gasto económico genera la producción puede ser muy grande.

En la Tabla 5.4 se observa más específico cual esquema de fusión fue el que obtuvo mejor resultado en cada serie de tiempo.

Tabla 5.4: Pronóstico con mejor resultado en cada serie de tiempo.

Serie de tiempo	Modelo	MSE
Aristeo-Mercado	Promedio (ANN + RF)	0.004670
El Fresno	Promedio (ANN + LR + RF)	0.009451
La Palma	Promedio (ANN + RF)	0.009179
La Piedad	Promedio (ANN + LR + RF)	0.006253
Cointzio	Promedio (ANN + <i>k</i> -NN + RF)	0.005278
Malpais	Promedio (ANN + RF)	0.005626
Patzcuaro	Promedio (ANN + RF)	0.000458

Implementar un esquema de fusión puede ofrecer un mejor resultado con respecto a un método ordinario, pero es probable que no en todos los casos se obtenga un buen resultado, es necesario aplicar esta metodología en otras áreas (bolsa de valores, medicina, agricultura, etc.) para determinar si el método es bueno; además, el proceso de entrenamiento es fundamental en estos procedimientos para obtener un buen resultado.

En las siguientes gráficas (Figuras 5.1 y 5.2) se muestran los resultados de las predicciones de series de tiempo tomadas de diferentes locaciones eólicas del estado de Michoacán. Estas gráficas son los resultados del mejor modelo de predicción utilizado en cada serie de tiempo, donde la línea continua representa los datos reales y la línea discontinua representa el pronóstico.

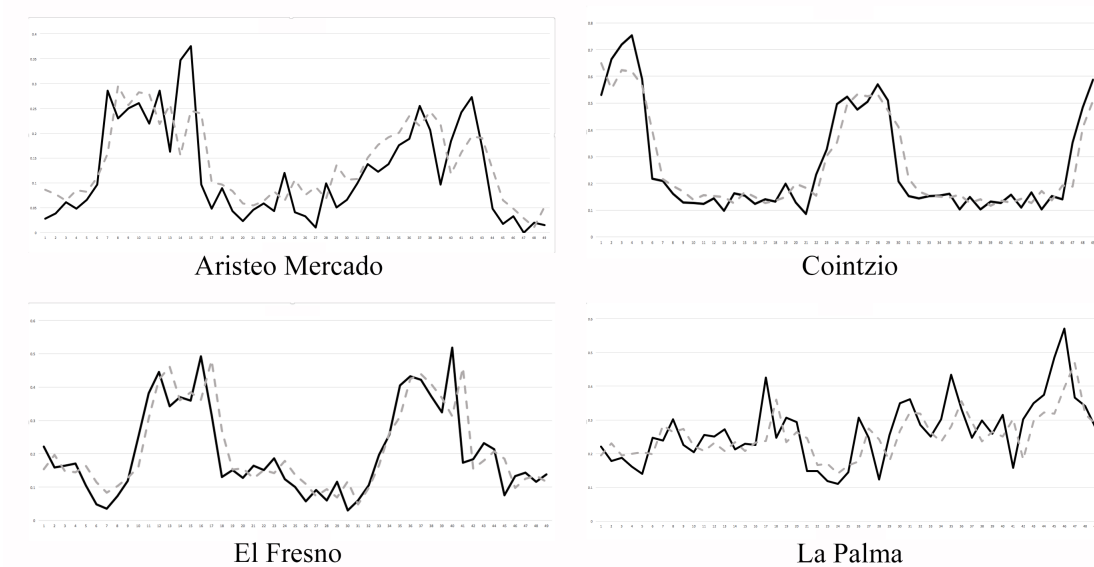


Figura 5.1: Resultados de los pronósticos implementando un esquema de votación de series de tiempo tomadas de estaciones eólicas de diferentes localidades del Michoacán.

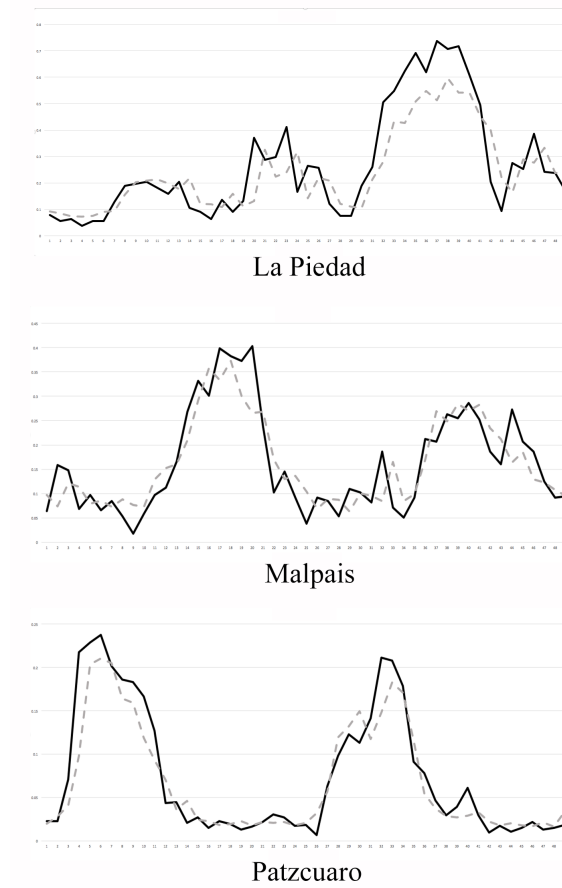


Figura 5.2: Resultados de los pronósticos implementando un esquema de votación de series de tiempo tomadas de estaciones eólicas de diferentes localidades del Michoacán.

Capítulo 6

Conclusiones

6.1. Conclusiones

Durante el desarrollo del trabajo de tesis se realizaron diferentes predicciones de series de tiempo de la velocidad del viento, aplicando un conjunto de métodos de predicción, tales como: redes neuronales, vecinos cercanos, regresión lineal y árboles aleatorios.

Con el resultado de cada uno de estos pronosticadores, se implementó un esquema de fusión con la finalidad de comprobar si aplicando una fusión de los datos pronosticados de las series de tiempo se obtendría un mejor resultado comparado con los resultados individuales de cada uno de los métodos de IA implementados.

En este esquema se aplicaron tres diferentes enfoques de fusión de datos (mediana, promedio y agrupamiento de datos), al obtener cada pronóstico de las series de tiempo se seleccionó la mejor fusión y se comparó el resultado con cada técnica de IA tomando como base la medida de aptitud MSE.

Al final del trabajo desarrollado, los mejores resultados se obtuvieron con la implementación del esquema de fusión en cada una de las series de tiempo utilizadas en los experimentos del trabajo de tesis, teniendo como conclusión de que implementando un esquema de fusión utilizando diferentes técnicas de IA, favorece a obtener mejores resultados en comparación con la aplicación de las técnicas individuales como: ANN, k -NN, LR y RF en series de tiempo de la velocidad del viento.

6.2. Aportaciones

Gracias a los resultados obtenidos con esta propuesta de fusión con diferentes métodos de IA es posible considerar el estudio más a fondo acerca de otras técnicas de fusión de datos, ya que han mostrado un alto desempeño en el pronóstico de series de tiempo.

Además se puede proponer el uso de otras técnicas de IA para aplicar la fusión de datos, por ejemplo, maquinas de soporte vectorial, y no solo implementar la fusión de datos con métodos de IA, sino también utilizar otros métodos de predicción como estadísticos, físicos, entre otros.

Otra aportación del trabajo de tesis es que se tiene una opción más de pronosticadores para realizar predicciones de series de tiempo de la velocidad del viento.

6.3. Trabajo futuro

Como trabajo futuro este trabajo servirá como referencia para realizar predicciones de series de tiempo fusionando diferentes métodos de predicción, no solamente de IA, se puede proponer el uso de otros métodos como estadísticos, físicos, de persistencia, entre otros.

Además, se pretende implementar el esquema de fusión con diferentes medidas o cálculos matemáticos, un ejemplo de ello sería aplicar el esquema de fusión mediante una ponderación de pesos a los pronosticadores utilizados en la fusión de datos.

Bibliografía

- Alomá, E. and Malaver, M. (2007). Análisis de los conceptos de energía, calor, trabajo y el teorema de carnot. *Enseñanza de las ciencias: revista de investigación y experiencias didácticas*, 25(3):387–400. 6, 7
- Baldi, P. and Brunak, S. (2001). *Bioinformatics: the machine learning approach*. MIT press. 15
- Barber, C., Bockhorst, J., and Roebber, P. (2010). Auto-regressive hmm inference with incomplete data for short-horizon wind forecasting. In *Advances in Neural Information Processing Systems*, pages 136–144. 4, 11
- Behl, R., Chhibar, R., Jain, S., Bahl, V., and El Bassam, N. (2012). Renewable energy sources and their applications. In *Proceedings of the International Conference on Renewable Energy for Institutes and Communities in Urban and Rural Settings*, pages 27–29. 8
- Boyle, G. (1996). *Renewable energy: power for a sustainable future*, volume 2. Oxford University Press. 3, 4, 8
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32. 17
- Brockwell, P. J. and Davis, R. A. (2016). *Introduction to time series and forecasting*. springer. 9
- Cao, J. and Cao, S. (2006). Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis. *Energy*, 31(15):3435–3445. 27

- Chang, W.-Y. (2013). Application of back propagation neural network for wind power generation forecasting. *International Journal of Digital Content Technology and Its Applications*, 7(4):502. 2
- Chang, W.-Y. (2014). A literature review of wind forecasting methods. *Journal of Power and Energy Engineering*, 2(04):161. 2, 11, 12, 22, 28
- Cheng, J. (2017). *Biomass to renewable energy processes*. CRC press. 7
- Cobo, B. and Batanero, C. (2000). La mediana en la educación secundaria obligatoria:¿ un concepto sencillo. *UNO*, 23:85–96. 19
- Cobo Merino, B. and Batanero, C. (2004). Significado de la media en los libros de texto de secundaria. *Enseñanza de las Ciencias*, 22(1):005–18. 19
- Doménech, J. L., Gil-Pérez, D., Martínez-Torregrosa, J., Gras, A., Guisasola, G., and Salinas, J. (2001). La enseñanza de la energía en la educación secundaria. un análisis crítico. *Revista de Enseñanza de la Física*, 14(1):45–60. 6
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231. 20
- Farmer, J. D. and Sidorowich, J. J. (1987). Predicting chaotic time series. *Physical review letters*, 59(8):845. 12
- Flores, J. J., Calderón, F., and Lara, J. O. C. (2015). Aprendizaje de modelos difusos para predicción de series de tiempo. *Revista Iztatl Computación*, (8):10. 24
- Ghosh, K., Ng, Y. S., and Srinivasan, R. (2011). Evaluation of decision fusion strategies for effective collaboration among heterogeneous fault diagnostic methods. *Computers & chemical engineering*, 35(2):342–355. 26
- Haque, A. U., Mandal, P., Meng, J., Srivastava, A. K., Tseng, T.-L., and Senjyu, T. (2012). A novel hybrid approach based on wavelet transform and fuzzy artmap network for pre-

- dicting wind farm power production. In *Industry Applications Society Annual Meeting (IAS), 2012 IEEE*, pages 1–8. IEEE. 26
- Hayashi, M. and Kermanshahi, B. (2001). Application of artificial neural network for wind speed prediction and determination of wind power generation output. In *Proceedings of ICEE*, pages 12–15. 3
- Hong, T. et al. (2014). Energy forecasting: Past, present, and future. *Foresight: The International Journal of Applied Forecasting*, (32):43–48. 4
- Kane, M. J., Price, N., Scotch, M., and Rabinowitz, P. (2014). Comparison of arima and random forest time series models for prediction of avian influenza h5n1 outbreaks. *BMC bioinformatics*, 15(1):276. 25
- Kirchgässner, G. and Wolters, J. (2007). *Introduction to modern time series analysis*. Springer Science & Business Media. 9
- Lei, M., Shiyang, L., Chuanwen, J., Hongling, L., and Yan, Z. (2009). A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews*, 13(4):915–920. 8, 12, 23
- Leven, R. and Koch, B. (1981). Chaotic behaviour of a parametrically excited damped pendulum. *Physics Letters A*, 86(2):71–74. 12
- Liaw, A., Wiener, M., et al. (2002). Classification and regression by randomforest. *R news*, 2(3):18–22. 17
- López, M. V. (2012). *Ingeniería de la energía eólica*, volume 5. Ediciones Marcombo. 9
- Maguire, L. P., Roche, B., McGinnity, T. M., and McDaid, L. (1998). Predicting a chaotic time series using a fuzzy neural network. *Information Sciences*, 112(1-4):125–136. 12
- Montgomery, D. C., Jennings, C. L., and Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons. x, x, 3, 10, 44
- Montgomery, D. C., Peck, E. A., and Vining, G. G. (2012). *Introduction to linear regression analysis*, volume 821. John Wiley & Sons. 16

- Moragues, J. A. and Rapallini, A. T. (2004). Aspectos ambientales de la energía eólica. *Documento del Instituto Argentino de la Energía, Buenos Aires.* 9
- Negnevitsky, M. (2005). *Artificial intelligence: a guide to intelligent systems.* Pearson Education. x, x, 13, 33
- Observ, E. (2012). Worldwide electricity production from renewable energy sources. *Ninth inventory.* 1, 2, 7
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1):217–222. 18
- Pérez-Landazábal, M., Favieres, A., Manrique del Campo, M., and Varela, P. (1995). La energía como núcleo en el diseño curricular de la física. *Enseñanza de las Ciencias*, 13(1):055–65. 7
- Pértega Díaz, S. and Pita Fernández, S. (2000). Técnicas de regresión: Regresión lineal múltiple. *Cad Aten Prim*, 7:173–76. 16
- Piwko, D. and Jordan, G. (2010). The economic value of day-ahead wind forecasts for power grid operations. In *2010 UWIG Workshop on Wind Forecasting.* 4
- Rabinowitz, P. M., Galusha, D., Vegso, S., Michalove, J., Rinne, S., Scotch, M., and Kane, M. (2012). Comparison of human and animal surveillance data for h5n1 influenza a in egypt 2006–2011. *PloS one*, 7(9):e43851. 26
- Raol, J. R. (2015). *Data fusion mathematics: theory and practice.* CRC Press. 19
- Rodriguez, H., Flores, J., Puig, V., Morales, L., Guerra, A., and Calderon, F. (2017). Wind speed time series reconstruction using a hybrid neural genetic approach. In *IOP Conference Series: Earth and Environmental Science*, volume 93, page 012020. IOP Publishing. 2, 11, 25
- Seber, G. A. and Lee, A. J. (2012). *Linear regression analysis*, volume 329. John Wiley & Sons. 16

- Sfetsos, A. (2002). A novel approach for the forecasting of mean hourly wind speed time series. *Renewable energy*, 27(2):163–174. 25
- Shafer, G. (1992). Dempster-shafer theory. *Encyclopedia of artificial intelligence*, pages 330–331. 26
- Soman, S. S., Zareipour, H., Malik, O., and Mandal, P. (2010). A review of wind power and wind speed forecasting methods with different time horizons. In *North American power symposium (NAPS), 2010*, pages 1–8. IEEE. 18
- Sorjamaa, A. and Lendasse, A. (2006). Time series prediction using dirrec strategy. 15
- Wang, W.-C., Chau, K.-W., Cheng, C.-T., and Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of hydrology*, 374(3-4):294–306. 23
- Wolfram, S. et al. (1996). *Mathematica*. Cambridge university press Cambridge. 36, 38
- Wu, Y.-K. and Hong, J.-S. (2007). A literature review of wind forecasting technology in the world. In *Power Tech, 2007 IEEE Lausanne*, pages 504–509. IEEE. 12, 18, 22
- Yadav, A. and Sahu, K. (2017). Wind forecasting using artificial neural networks: A survey and taxonomy. *International Journal of Research In Science & Engineering*, 3. 12
- Yuan-Kang, W., Ching-Ying, L., Shao-Hong, T., and Yu, S.-N. (2010). Actual experience on the short-term wind power forecasting at penghu?from an island perspective. In *Power System Technology (POWERCON), 2010 International Conference on*, pages 1–8. IEEE. 2, 14