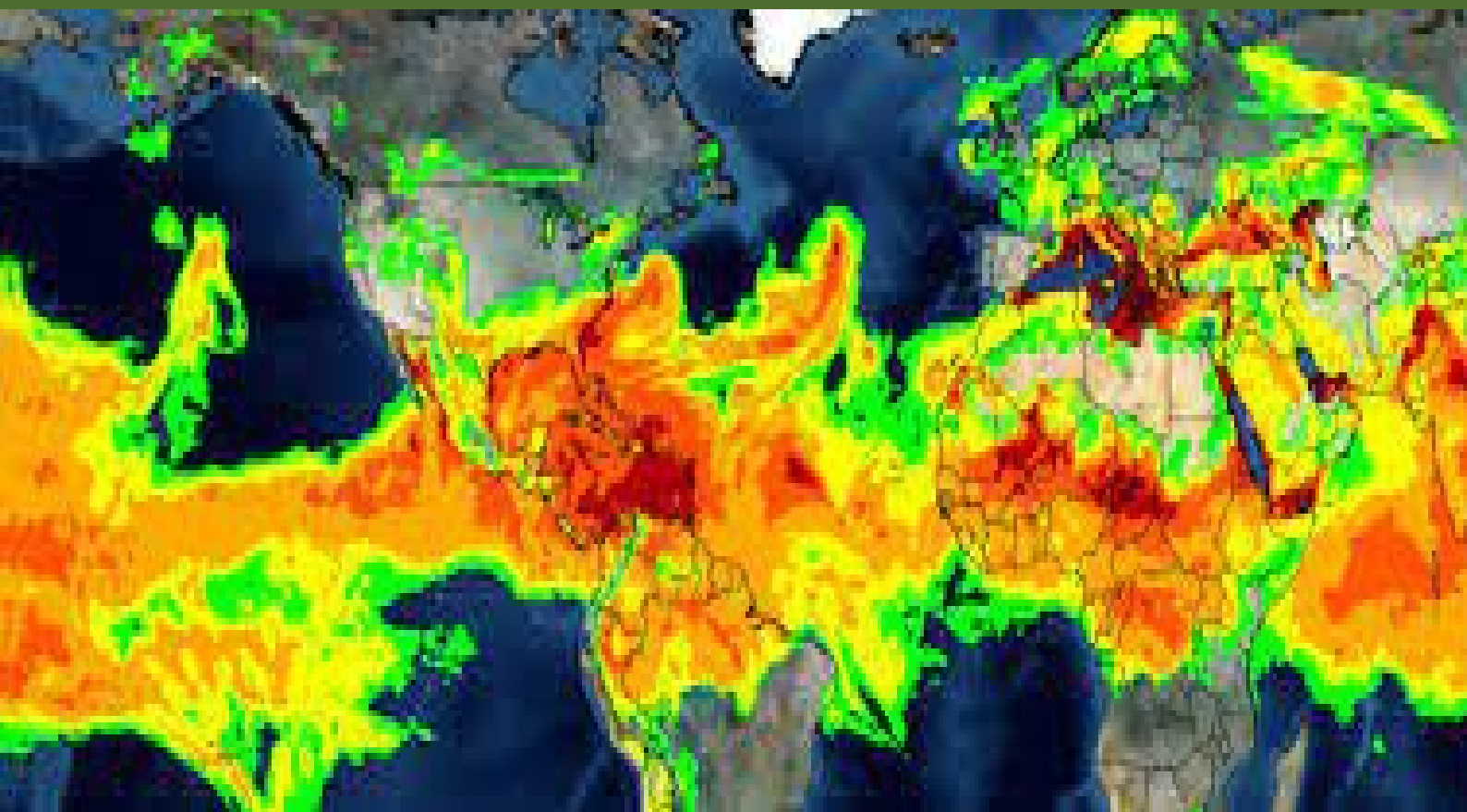


# Data mining para determinar patrones del comportamiento de datos meteorológicos

Jorge Huere-Peña  
Jose Gave-Chagua  
Russbelt Yaulilahua-Huacho

William Salas-Contreras  
Teresa Gonzales  
Jose Ayuque-Rojas



**DOI: 10.35622/inudi.b.066**

EDITADA POR  
INSTITUTO  
UNIVERSITARIO  
DE INNOVACIÓN CIENCIA  
Y TECNOLOGÍA INUDI PERÚ





# **Data mining para determinar patrones del comportamiento de datos meteorológicos**

**DOI: <https://doi.org/10.35622/inudi.b.066>**

**Jorge Huere-Peña**

**<https://orcid.org/0000-0002-3114-8134>  
[jorge.huere@unh.edu.pe](mailto:jorge.huere@unh.edu.pe)**

**Jose Gave-Chagua**

**<https://orcid.org/0000-0002-0434-3663>  
[jose.gave@unh.edu.pe](mailto:jose.gave@unh.edu.pe)**

**Russbelt Yaulilahua-Huacho**

**<https://orcid.org/0000-0002-7007-3059>  
[russbeltyauli24@gmail.com](mailto:russbeltyauli24@gmail.com)**

**William Salas-Contreras**

**<https://orcid.org/0000-0001-7664-3000>  
[william.salas@unh.edu.pe](mailto:william.salas@unh.edu.pe)**

**Teresa Gonzales**

**<https://orcid.org/0000-0003-2945-3874>  
[teresa.gonzales@unh.edu.pe](mailto:teresa.gonzales@unh.edu.pe)**

**Jose Ayuque-Rojas**

**<https://orcid.org/0000-0002-8735-4105>  
[jayuque@unah.edu.pe](mailto:jayuque@unah.edu.pe)**

**Instituto Universitario  
de Innovación Ciencia y Tecnología Inudi Perú**

Data mining para determinar patrones del comportamiento de datos meteorológicos

Jorge Luis Huere Peña  
Jose Luis Gave Chagua  
Russbelt Yaulilahua Huacho  
William Herminio Salas Contreras  
Teresa Jesus Gonzales Huamán  
Jose Carlos Ayuque Rojas  
(Autores)

ISBN: 978-612-5069-55-9 (PDF)

Hecho el depósito legal en la Biblioteca Nacional del Perú N° 2022-12959

DOI: <https://doi.org/10.35622/inudi.b.066>

Editado por Instituto Universitario de Innovación Ciencia y Tecnología Inudi Perú S.A.C  
Urb. Ciudad Jardín Mz. B3 Lt. 2, Puno – Perú

RUC: 20608044818

Email: [editorial@inudi.edu.pe](mailto:editorial@inudi.edu.pe)

Teléfono: +51 973668341

Sitio web: <https://editorial.inudi.edu.pe>

Primera edición digital

Puno, diciembre de 2022

Libro electrónico disponible en

<https://doi.org/10.35622/inudi.b.066>

**Editores:**

Wilson Sucari / Patty Aza / Antonio Flores

*Las opiniones expuestas en este libro es de exclusiva responsabilidad del autor/a y no necesariamente reflejan la posición de la editorial.*

*Publicación sometida a evaluación de pares académicos (Peer Review Doubled Blinded)*

Publicado en Perú / *Posted in Peru*



Esta obra está bajo una licencia internacional Creative Commons Atribución 4.0.





## Contenido

SINOPSIS .....	9
ABSTRACT .....	10
INTRODUCCIÓN .....	11
CAPÍTULO I .....	13
CARACTERIZACIÓN DE LA INVESTIGACIÓN Y MARCO METODOLÓGICO	13
1.1. Descripción de problema.....	13
1.2 Enunciado de problema .....	14
1.2.1 Enunciado general del problema .....	14
1.2.2 Enunciados específicos del problema.....	14
1.3 Objetivo de la investigación .....	14
1.4 Método, diseño y tipo de investigación .....	14
1.5 Cuadro de variables, temas o unidades de investigación .....	16
1.6 Técnicas e instrumentos de investigación .....	16
1.7 Procedimientos de investigación .....	18
1.8 Consideraciones éticas .....	18
CAPITULO II.....	19
MARCO TEÓRICO .....	19
2.1 Bases teóricas .....	19
2.1.1. Cambio climático y variabilidad climática .....	19
2.1.2. Meteorología, tiempo meteorológico y clima .....	19
2.1.3. Patrones de comportamiento de datos meteorológicos .....	20
2.2 Marco conceptual .....	21
2.3 Marco filosófico .....	36
CAPÍTULO III .....	41
RESULTADOS, DISCUSIONES Y CONCLUSIONES.....	41
3.1. Exposición resultados .....	41
3.1.1. Resultado 1 .....	41
3.1.2. Resultado 2 .....	42
3.1.3. Resultado 3.....	42
3.1.4. Resultado 4 .....	43
3.1.5. Resultado 5.....	45
3.1.5. Resultado 5.....	46
3.1.6. Resultado 6 .....	48

3.1.7. Resultado 7 .....	48
3.1.8. Resultado 8 .....	50
3.1.9. Resultado 9 .....	51
3.2. Discusión .....	57
3.3. Conclusiones .....	60
REFERENCIAS .....	61
ANEXOS .....	63





## SINOPSIS

El libro es una adaptación de una investigación presentada a la Universidad Nacional de Huancavelica, que tuvo como objetivo determinar patrones de comportamiento de datos obtenidos mediante éstas técnicas, de las variables meteorológicas en la ciudad de Huancavelica (Perú) como son: la temperatura ambiental, presión atmosférica, humedad atmosférica, velocidad del viento, radiación solar, radiación ultra violeta y precipitación pluvial utilizando para ello una estación meteorológica automatizada de la compañía Weather Link, Marca DAVIS, Modelo Vantage Pro y una consola para el almacenamiento de datos Vantage Pro en texto plano y que posteriormente fueron procesados, descritos y analizados usando el software SPSS Statistical y WRPLOT en el caso particular de la variable dirección del viento y para la determinación de comportamientos y patrones se usó la metodología CRISP-DM. Los resultados obtenidos fueron clúster de las variables meteorológicas con algoritmos de aprendizaje no supervisado y predicciones de la variable precipitación pluvial con algoritmos de aprendizaje supervisados obteniendo 84,9% de probabilidades de éxito en el pronóstico y en el caso de los clúster grupos de cuatro y diez significativamente diferentes.

**Palabras clave:** variables meteorológicas, patrones meteorológicos, comportamiento meteorológico, data mining.

## ABSTRACT

The book is an adaptation of an investigation presented to the National University of Huancavelica, which aimed to determine patterns of behavior of data obtained through these techniques, of the meteorological variables in the city of Huancavelica (Peru) such as: ambient temperature, atmospheric pressure, atmospheric humidity, wind speed, solar radiation, ultraviolet radiation and rainfall using an automated weather station from the Weather Link company, DAVIS brand, Vantage Pro model and a Vantage Pro data storage console in text plane and that were later processed, described and analyzed using the SPSS Statistical and WRPLOT software in the particular case of the wind direction variable and for the determination of behaviors and patterns the CRISP-DM methodology was used. The results obtained were a cluster of meteorological variables with unsupervised learning algorithms and predictions of the rainfall variable with supervised learning algorithms, obtaining 84.9% probability of success in the forecast and in the case of the clusters groups of four and ten significantly different.

**Keywords:** meteorological variables, meteorological patterns, meteorological behavior, data mining.

## INTRODUCCIÓN

El cambio del clima se ha convertido en un factor a tener en cuenta por los países para adoptar medidas de adaptación que reduzcan los efectos negativos al ambiente, infraestructura, salud, economía entre otros (Magaña Rueda & Gay García, 2014), es por eso que el estudio de los meteoros de la atmósfera es importante para lograr estos resultados debido a eso se realizan estudios a nivel mundial para predecir los fenómenos meteorológicos y con el ello los cambio del clima a nivel nacional y mundial (Zúñiga López & Crespo del Arco, 2010).

Es así que el estudio de la variación de los meteoros que se presentan en la atmósfera terrestre en periodos determinados es de utilidad para los estudios del clima a nivel regional y local (Páez Martínez, Espitia Barrera, & Cárdenas Castro, 2013), como lo han sido desde tiempos pasados en los que su importancia radica en la aplicación que se le encuentra en diversas temáticas como el de la agricultura, la salud, etc. (Alberola Romá, 2015), además de que los estudios son realizados a nivel rural y urbano siendo necesario estudiar la data meteorológica para los estudios de la variación meteorológica que en cada región es característica (De Schiller, Evans, & Katschner, 2001).

La variación meteorológica es una característica que depende de factores como los meteoros atmosféricos, factores geográficos, temporalidad así también las actividades antropogénicas las que en el último siglo se han convertido en aspectos que han intensificado el cambio climático, las investigaciones se centrar en el estudio de la variación en periodos de tiempo de cada uno de los meteoros y la interpretación de la variación presente en periodos determinados de tiempo aunque la limitada cantidad de estaciones meteorológicas y el periodo de registro es muchas veces una limitante (Pezoa Gutiérrez, 2003), también es de interés por el efecto que tienen en el comportamiento de factores biológicos (Galán, Infante, Ruiz de Clavijo, & Domínguez, 1988).

La estimación del peligro de incendios a partir de la teledetección y las variables meteorológicas que han resultado útiles en muchos países (Yevra Álvarez, De Santis, & Chuvieco, 2005), Las investigaciones de la variación de la precipitación también son de vital importancia por la problemática de disponibilidad de agua

en el planeta es así que estudiar la variación de las lluvias es considerado como estudios indispensables en los países (Campo, Ramos, & Zapperi, n.d.).

El estudio de la disponibilidad de agua es necesario para cuantificar el ingreso y salida de agua de los ecosistemas y su uso de acuerdo a las actividades humanas y las necesidades de los componentes de estos espacios territoriales conocidos como cuencas hidrográficas a este tipo de estudios se le conoce como el balance hídrico de una cuenca por eso este tipo de investigaciones son realizadas en muchos países para afrontar la disponibilidad de agua de acuerdo a la demanda de este recurso (Marini & Piccolo, 2000).

Para realizar estos estudios se requiere los datos de precipitación pluvial y de temperatura ambiental con el que se calcula la evapotranspiración con el cual se tiene los elementos necesarios para realizar el balance hídrico el que es realizado de un año o periodo determinado en el que se realiza el balance del agua que se produce en una cuenca la que en diversas áreas del territorio es diferente así como en periodos diferentes (Echeverría, Huber, & Taberlet, 2007), con esta información y otras que se pueda obtener a partir de técnicas no convencionales como el data Mining se planifica las diversas actividades de una cuenca hidrográfica en función a la demanda de agua, el balance hídrico y otros patrones de comportamiento que se pueda encontrar, así también el uso de las imágenes satelitales es actualmente una herramienta para investigar datos meteorológicos en diferentes escenarios y ámbitos de estudio (Jobbagy Gampel, Acosta, & Nosetto, 2016).

# CAPÍTULO I

## CARACTERIZACIÓN DE LA INVESTIGACIÓN Y MARCO METODOLÓGICO

### 1.1. Descripción de problema

Actualmente, debido al desarrollo desmedido de las tecnologías de la información y la comunicación, vivimos en una situación de crisis y en una situación de cambios cada vez más rápidos, en la era de la información, digital o informática, de la que nace una sociedad de la información, donde los aspectos sociales, económicas, políticas, religiosas, educativas, etc. Eliminaron las barreras de tiempo y espacio debido que resultan de la creación de una comunicación ubicua.

Esta nueva sociedad de la información y los recursos informáticos nuevos y existentes cuentan hoy con datos (Big Data) cada vez más diversos: geoespaciales, bancarios, educativos, empresariales, meteorológicos, etc. cuya complejidad por tamaño y ritmo de crecimiento lo dificulta. Recopile procese y gestione con las herramientas tradicionales de Big Data.

Esta investigación nos permite encontrar conexiones inherentes entre las nuevas técnicas de procesamiento de datos aplicadas a los datos meteorológicos, como la minería de datos Data Mining, que tiene como objetivo descubrir nuevos patrones de comportamiento de estas técnicas y los algoritmos informáticos que contienen. Puede ayudar a tomar decisiones más acertadas que apoyen la gestión ambiental y conocer sus patrones y tendencias.

Las técnicas de procesamiento de datos convencionales no son las únicas fuentes de toma de decisiones, la implementación de nuevas técnicas ayudará a los gestores ambientales y personas involucradas en el área a realizar de manera más eficiente su labor, siempre que ellas se encuentren plenamente sustentadas y comprobadas científicamente, mediante pruebas y determinación de patrones de comportamiento de datos meteorológicos, para su inferencia en tiempo y espacio de manera más adecuada y alto grado de precisión posible.

La gran cantidad de datos meteorológicos existentes y crecientes crea la necesidad de desarrollar métodos y técnicas modernas que permitan procesar, analizar y determinar el comportamiento de los datos meteorológicos para tomar

las decisiones más acertadas y prevenir riesgos en función del tipo de entorno. Incertidumbre y probabilidad de ocurrencia cuando estos métodos computacionales hacen predicciones.

## 1.2 Enunciado de problema

### 1.2.1 Enunciado general del problema

¿Cuáles son los patrones de comportamiento obtenidos mediante técnicas de Data Mining, de datos asociados a las variables meteorológicas en la ciudad de Huancavelica, 2018 - 2019?

### 1.2.2 Enunciados específicos del problema

¿Existen diferencias significativas entre clúster formados mediante técnicas de Data Mining, a partir de valores meteorológicos diarios en la ciudad de Huancavelica, 2018 - 2019?

¿Existe diferencia significativa entre las variables meteorológicas procesados mediante técnicas de Data Mining y los diferentes meses del año en la ciudad de Huancavelica, 2018 - 2019?

## 1.3 Objetivo de la investigación

### **Objetivo general**

Determinar patrones de comportamiento de datos obtenidos mediante técnicas de Data Mining, asociados a las variables meteorológicas en la ciudad de Huancavelica, 2018 – 2019.

### **Objetivo específico**

Determinar si existen diferencias significativas entre clúster formados mediante técnicas de Data Mining, a partir de valores meteorológicos diarios en la ciudad de Huancavelica, 2018 – 2019.

Determinar si existe diferencia significativa entre las variables meteorológicas procesados mediante técnicas de Data Mining y los diferentes meses del año en la ciudad de Huancavelica, 2018 – 2019.

## 1.4 Método, diseño y tipo de investigación

Los métodos de investigación en el trabajo de investigación serán métodos longitudinales y series de tiempo, ya que se utilizarán datos meteorológicos

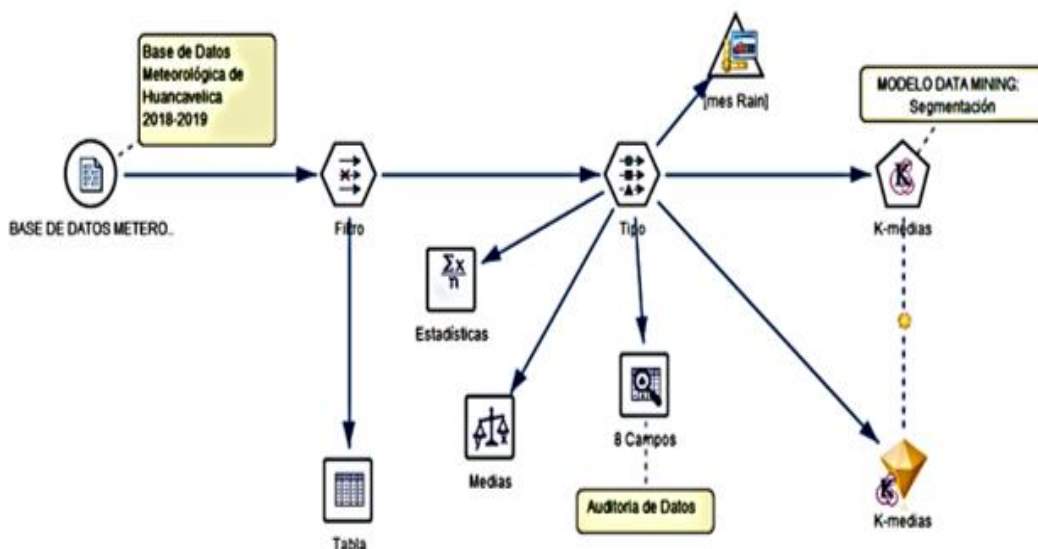
históricos para derivar modelos de comportamiento a partir de los datos meteorológicos de la ciudad de Huancavelica para su posterior, desarrollo y una adecuada gestión ambiental. Los datos de las variables meteorológicas se manipularán para revelar patrones y describir el comportamiento oculto de las variables en estudio utilizando técnicas de minería de datos *Data Mining*.

## Diseño

En este estudio se modeló el diseño con el software SPSS Modeler utilizado, partiendo de la base de datos recolectada por AWS, de los filtros utilizados en la base de datos, el preprocesamiento de datos, el modelo de algorítmico utilizado(K-Means) para el modelo propuesto en el estudio se encontró como se muestra en la Figura 5.

## Figura 1

*Diseño de la investigación mediante SPSS Modeler para determinar Patrones Meteorológicos*



## Tipo

Basado en el estudio actual de Sánchez y Reyes (1998) este método consiste en describir, analizar e interpretar sistemáticamente un conjunto de hechos relación con otras variables que están ocurriendo actualmente. La investigación propuesta en base a las características que presenta es de tipo **Descriptivo**, ya que se examina para describir las variables única del estudio para comprender el comportamiento de una realidad específica utilizando técnicas de minería de



datos *Data Mining*. Asimismo, el nivel de investigación es **Básico**, porque se utilizando técnicas de minería de datos *Data Mining* para derivar patrones de comportamiento a partir de datos meteorológicos para luego realizar la gestión ambiental relevante y la toma de decisiones en la ciudad de Huancavelica con base en el comportamiento meteorológicas.

### 1.5 Cuadro de variables, temas o unidades de investigación

**Tabla 1**

*Operacionalización de las variables meteorológicas de la ciudad de Huancavelica.*

VARIABLE	DIMENSION	INDICADORES	INSTRUMENTO	UNIDAD DE MEDIDA
Variables meteorológicas en la ciudad de Huancavelica 2018 -2019	Temperatura Ambiental	Temperatura promedio de 24Hr.		°C
	Precipitación Pluvial	Precipitación acumulada diaria		mm
	Humedad relativa	Humedad porcentual promedio en 24Hr.		%
	Presión atmosférica	Presión atmosférica promedio de 24Hr.	Estación Meteorológica Automatizada Weather Link Marca DAVIS	Mm de Hg
	Radiación Solar	Radiación acumulada 24Hr.	Modelo Vantage Pro y una consola Vantage Pro	W/m <sup>2</sup>
	Radiación Ultravioleta	Índice de radiación UV promedio de 24Hr.		UV Index
	Velocidad del viento	Velocidad promedio de 24Hr.		m/s

### 1.6 Técnicas e instrumentos de investigación

#### Técnicas

La técnica utilizada en este estudio de investigación es la observación utilizando instrumentos implementados en estaciones meteorológicas automáticas, como el Instituto Nacional de Meteorología e Hidrología del Perú - SENAMHI.

#### Los instrumentos

Los instrumentos y equipos que sirven para diversas variables meteorológicas, los cuales pueden ser manuales o convencionales, luego automáticas, cada día

más precisos, los otros son los más eficientes, esta función se puede realizar día a día pronosticando. Mas rápido y más preciso. Los instrumentos de recolección de datos de variables de medición utilizados en este estudio son la estación meteorológica automática Weather Link, marca DAVIS, modelo Vantage Pro y una consola Vantage Pro, como se muestra en la Figura 6, y están equipados con los mismos sensores para registrar varios parámetros meteorológicos ambientales. Ejemplo: temperatura ambiente (promedio, máxima, mínima e interna), presión atmosférica, humedad, velocidad del viento, dirección del viento, radiación solar, radiación ultra violeta y precipitación pluvial, etc., todo en forma automatizada y digital para su posterior procesamiento y análisis.

## Figura 2

*Estación Meteorológica Automatizada Weather Link, Marca DAVIS, Modelo Vantage Pro*



Se utilizan estaciones meteorológica automática en lugar de las estaciones meteorológica son más precisas y no requieren de grabadores humano, evitando así riesgos en la toma de datos, esta estación meteorológica automática proporciona temperatura ambiente, presión atmosférica, humedad atmosférica, velocidad del viento, radiación solar, radiación ultra violeta y precipitación pluvial, las mismas que fueron registradas en intervalos de una hora, los datos luego se descargaron en texto plano desde la consola Vantage Pro a microcomputadoras personales, es un año, hoja de cálculo electrónico.

## 1.7 Procedimientos de investigación

En la técnicas de procesamiento y análisis de datos, los datos importados desde consola Vantage Pro en formato de texto plano separado por comas CSV-UTF-8 se compilaron inicialmente para su análisis y cálculos descriptivos evalúan la tendencia central ( media aritmética, mediana y moda), medidas de dispersión ( varianza, desviación estándar y coeficiente de variabilidad) de estadística relevantes para describir los datos obtenidas de forma original y convencional de variable de velocidad y dirección del viento y diseño de gráficos de rosa de los vientos anual y mensuales utilizando WRPLOT View - Lakes Environmental Software – Wind Rose for Meteorological Data, análisis de datos y determinación de patrones usando técnicas de data mining se usó IBM SPSS Modeler software y para las pruebas de hipótesis se usó IBM SPSS V.25 Statistics (Statistical Package for the Social Sciences), como se muestra en la Figura 7.

### Figura 3

*Software usado para la descripción, análisis y Data Mining de los datos meteorológicos*



## 1.8 Consideraciones éticas

- Los docentes seleccionados como muestra de estudio tenían conocimiento de los alcances de la investigación, por lo cual hubo una participación informada.
- La identidad de cada uno de los docentes se mantiene en el anonimato para no invadir su intimidad profesional.
- Las fuentes consultadas fueron citadas y referenciadas según normas APA.

## CAPITULO II

### MARCO TEÓRICO

#### 2.1 Bases teóricas

##### 2.1.1. Cambio climático y variabilidad climática

El cambio climático se define como un cambio permanente y significativa en el registro climático a escala regional y global y en varios parámetros meteorológicos. La causa pueden ser el cambio climático tanto natural como provocado por el hombre, cambios en la composición química de la atmósfera provocados por un aumento global de los Gases de Efecto Invernadero (GEI). Se caracteriza por un incremento general y gradual de la temperatura, cambios en los comportamientos de las precipitaciones e incremento de eventos extremos (Serrano et al., 2012).

La variabilidad climática consiste en aquellos cambios sobre el clima que dependen de eventos climáticos extremos por encima de promedios estándares. Los fenómenos que producen estos contrastes pueden ser frentes fríos altamente organizados, células madre en calma, huracanes, perturbaciones tropicales y células con niveles de humedad desproporcionadamente altos (Serrano et al., 2012).

##### 2.1.2. Meteorología, tiempo meteorológico y clima

Meteorología, tiempo o tiempo meteorología y clima, parece referirse al mismo evento, pero no es así. La **meteorología** es el estudio de la atmósfera y de los mecanismos que provocan cambios en ella. El término **tiempo meteorológico** o simplemente clima, se refiere al mecanismo de cambio estudiado en cortos periodos de tiempo. En la escala de la vida humana, hablamos de un momento determinado, tiempo, estado de atmósfera en un momento determinado y lo que significa, por ejemplo, el tiempo en este caso suele referirse a la luz, la lluvia, los factores del aire y los factores eléctricos, básicamente. Estos términos incluyen hidrometeoros, lluvia, nieve, granizo, entre otros. Por otro lado, el concepto de **clima** es esencial se estudia variabilidad del tiempo a lo largo de periodo de tiempo meteorológico a largo plazo (multianuales) (Guerra, 2017).

### 2.1.3. Patrones de comportamiento de datos meteorológicos

En los últimos años, se han utilizado datos meteorológicos de alta resolución para desarrollar modelos climáticos y modelos de comportamiento que se han utilizado para estudiar los cambios a largo plazo particular el cambio climático actual. Sin embargo, hay que ser cuidadosos en este sentido: el clima es un promedio estadístico a largo plazo de los datos meteorológicos obtenidos en estaciones meteorológicas ubicadas en un área determinada y que tiene características similares que definen un clima determinado. Se hace en todos los climas en todo el mundo. Sin embargo, estos tipos de clima no pueden resumir en modelos o patrones de comportamiento específicos, ya que sus cambios a largo plazo de los mismos deben ser obtenidas a posteriori de dichas variaciones producidas a largo plazo.

En otras palabras: la información meteorológica obtenida de muchas estaciones meteorológicas de todo el mundo se usa de manera inductiva, para establecer las características del clima y sus variaciones en toda la tierra, y una vez obtenida, podemos estudiar los cambios climáticos ocurridos en el pasado hasta el momento en el que se analizan, pero no podríamos usar esta información hacia el futuro, porque la meteorología y la climatología trabajan a diferentes escalas, solo se usará como una probabilidad o un posible patrón de comportamiento.

### **Algoritmos computacionales y procesamiento de datos meteorológicos**

Debido a la gran cantidad de datos meteorológicos obtenidos por diferentes sistemas de observación y fuentes meteorológicas, su gestión y análisis es un proceso complejo y no lineal a fin de obtener patrones de comportamiento. Esta complejidad para obtener patrones de comportamiento se puede solucionar aplicando las técnicas de aprendizaje automático con diferentes variables de entrada meteorológicas, usando algoritmos computacionales y técnicas estadísticas tales como los modelos de predicción y análisis de comportamiento k-Nearest Neighbour (kNN), Artificial Neural Networks (ANN), Adaptive Boosting (AdaBoost), Nature-Inspired Optimization Algorithms, Quick Cluster (k-medias)), Decision Tree Algorithm, Time series algorithms, entre otros para analizar, encontrar, descubrir y comprender patrones de comportamiento meteorológico con diferentes escenarios de datos meteorológicos disponibles

como: temperatura del aire (escenario 1), temperatura del aire y radiación solar (escenario 2), temperatura del aire, radiación solar y velocidad del viento (escenario 3), temperatura del aire, radiación solar, velocidad del viento y humedad relativa (escenarios 4) dirección del viento, velocidad del viento y humedad relativa mediante rosa de viento (escenarios 5) (Seda & Todorovic, 2020; Sujay et al., 2019; Aragón et al., 2019).

## 2.2 Marco conceptual

### **Aspectos generales**

#### **Meteorología**

La Meteorología se define como la ciencia encargada del estudio de la atmósfera, sus propiedades y los fenómenos que en ella ocurren (llamados meteoritos). El estudio de la atmósfera se basa en el conocimiento de una serie de variables cuantitativas o meteorológicas, como la temperatura, la presión atmosférica, la radiación solar, la velocidad del viento y otras variables que varían en el espacio y el tiempo (Fundación Española para la Ciencia y la Tecnología, 2004).

#### **Descripción de la atmósfera**

La atmósfera terrestre es la capa de aire que rodea la Tierra que es atraída por la gravedad y consiste en masas de aire y una mezcla de gases. Los principales gases que componen el aire terrestre son: nitrógeno ( $N_2$ ) 78.084%, oxígeno ( $O_2$ ) 20.95%, argón (Ar) 0.93%, anhídrido carbónico ( $CO_2$ ) 0.03%, Neón (Ne) 0.02%, Helio (He) 0.0005%, para Metano ( $CH_4$ ) 0.001%, para Criptón (Kr) 0.001% y vapor de agua, este último importante para la meteorología (Oldani, 2020). Es importante aclarar que la concentración de estos gases varía con la altitud. Cuando hay una transición entre dos masas de aire de diferentes densidad y temperatura, los llamados frentes no solo producen lluvia, sino que también provocan cambios en la presión atmosférica, temperatura, viento y otros fenómenos (Guerra, 2017) (Fundación Española para la Ciencia y la Tecnología, 2004).

## **Organización Meteorológica Mundial (OMM)**

El objetivo principal del Sistema Mundial de Procesamientos de Datos y Pronósticos (SMPDP) es preparar y poner a disposición de los Miembros, productos de análisis y pronóstico meteorológicos de manera más rentable y su estructura organizativa está conformado por los Centros Meteorológicos Mundiales (CMM), los Centros Meteorológicos Regionales Especializados (CMRE) y los Centros Meteorológicos Nacionales (CMN) de cada país, en algunos casos forman un único centro y es responsable del desempeño simultáneo de las funciones de estos centros (Organización Meteorológica Mundial, 2010)

## **Fuentes de información meteorológica**

Las fuentes de información meteorológica permiten analizar variables y fenómenos meteorológicos que se encuentran en miles de estaciones meteorológicas ubicadas en lugares específicos de la superficie terrestre en posiciones fijas, pero también sobre el mar (barcos, buques y otros) ya diferentes alturas de la atmósfera (satélites y aviones) aprovechando que todos tienen instrumentos meteorológicos a bordo.

El uso de la información obtenida de todas estas fuentes de observación es muy amplio y variado: desde el mero registro temporal en determinadas estaciones la elaboración pronóstico meteorológicas. En cualquier caso, los centros meteorológicos agregan datos por región, los procesan, comprueban su calidad y los comparten con los usuarios que puedan necesitarlos para las investigaciones y estudio de la atmósfera.

## **Calidad y validación de datos meteorológicos**

Los datos meteorológicos, independientemente de la fuente de su información, deben ser considerados cualitativos, lo cual responden a necesidades específicos y hacen uso de esta información de forma oculta. Para ello, es necesario asegurarse de que los datos cumplen las condiciones necesarias (en términos de certeza, resolución, continuidad, homogeneidad, representatividad, puntualidad, formato, etc.), su calidad debe ser conocida (dentro del rango) y verificable. La Organización Mundial de Meteorología se basa en la norma de la *International*

*Organization for Standardization ISO 9000: Sistema de Gestión de la Calidad* (World Meteorological Organization, 2017).

Los principios y los fundamentos en que se basan la mayoría de los tests que se aplican a datos registrados en una estación como procedimientos para validación de datos sobre diversas variables meteorológicos, existen cinco: rangos o intervalos, consistencia temporal, persistencia, consistencia interna y consistencia espacial (Estévez, 2008).

El tests de rangos determina si las observaciones están dentro de un rango predeterminado que puede ser fijo o dinámico, tal como se muestra en la Tabla 1.

**Tabla 2**

*Límites físicos de diferentes variables meteorológicas.*

<b>Variable</b>	<b>Unidad</b>	<b>Rango</b>
Temperatura del aire	°C	-35/55
Humedad Relativa del aire	%	0/100
Velocidad de viento	m/s	0/75
Dirección de viento	grados	0/360
Presión	hPa	700/1080
Radiación Solar Global	W/m <sup>2</sup>	-1/1400
Precipitación en 10 min	mm	0/50

La consistencia temporal es un procedimiento para comprobar si la diferencia entre medidas meteorológicas sucesivas excede un valor determinado, en cuyo caso medidas deben ser sospechosas, (OMM) solo recomienda un rango de tolerancia para las variables temperatura  $T$ =temperatura del aire,  $T_{dew}$ =temperatura del punto de rocío y presión, tal como se muestra en la Tabla 2, donde  $dt$  es la diferencia de tiempo entre los dos datos consecutivos y  $TOL$  es la tolerancia máxima permitida (World Meteorological Organization, 1993).



### Tabla 3

*Tolerancias propuestas en función del tiempo transcurrido entre medidas consecutivas*

	<b>dt=1 hora</b>	<b>dt=2 hora</b>	<b>dt=3 hora</b>	<b>dt=6 hora</b>	<b>dt=12 hora</b>
<b>T TOL</b>	4 °C	7 °C	9 °C	15 °C	25 °C
<b><math>T_{dew}</math> TOL</b>	4 °C	6 °C	8 °C	12 °C	20 °C
<b>Presión TOL</b>	3 hPa	6 hPa	9 hPa	18 hPa	36 hPa

La persistencia es un procedimiento basado en la consistencia temporal. En lugar de evaluar grandes cambios o diferencias entre observaciones sucesivas, ahora se comprueba la escasa o nula variabilidad de dichos registros. La consistencia interna por otro lado, es un procedimiento basado en la verificación de la coherencia física o climatológica de cada variable observada o también de la consistencia entre variables. La consistencia espacial es un algoritmo utilizado para detectar posibles errores al comparar los datos de una estación con los valores correspondientes a estaciones vecinas (World Meteorological Organization, 1993).

### **Variables meteorológicas y patrones de comportamiento**

#### **a. Variables meteorológicas, concepto, instrumento y unidades de medida.**

La precipitación se define como el producto líquido o sólido de la condensación del vapor de agua que cae de las nubes o el aire y se deposita por gravedad en el suelo, incluida la precipitación de lluvia, granizo, nieve, rocío, cencellada blanca, escarcha y niebla. Un instrumento que se utiliza para medir la cantidad de lluvia en un lugar específico y en un periodo de tiempo se llama pluviómetro. La precipitación se mide en milímetros de agua o litros que caen sobre la superficie unidad  $l/m^2$  (volumen/área) o en  $kg/m^2$  (masa/área) para la precipitación líquida (World Meteorological Organization, 2017).

La Humedad es un término que se usa a menudo para describir la cantidad de vapor de agua presente en el aire. Los medidores de humedad están disponibles en varios tamaños. La relación de mezcla, que es la relación entre la masa de vapor de agua y la masa de aire seco, o la humedad específica, que es la relación

entre la masa de vapor de agua y la masa de aire húmedo. Humedad relativa, expresada como la relación entre la presión de vapor observada y la tensión del vapor saturante con respecto al agua a la misma temperatura y presión, expresada como porcentaje (WMO, 2015).

El viento es el movimiento del aire en una determinada dirección y velocidad. La Dirección del viento indica de dónde viene o sopla el viento y es un valor derivado de una dirección de viento promedio de 10 minutos, su unidad de medición se representa en grados Dextrorsum (giro en sentido de las manecillas del reloj) donde  $0^\circ$  es el norte verdadero y el instrumento para medir la dirección del viento se utilizan las veletas (Fundación Española para la Ciencia y la Tecnología, 2004; Guerra, 2017).

La velocidad del viento es la medida aritmética de las velocidades medidas durante un periodo de un lapso de 10 minutos, su unidad de medición para estaciones meteorológicas automáticas (EMA) es km/h, en estaciones Meteorológicas Sinópticas (ESIME) en m/s. El instrumento más utilizado para medir la velocidad del viento es el anemómetro de cazoletas (Fundación Española para la Ciencia y la Tecnología, 2004; Mendoza et al., 2020).

La temperatura es una magnitud relacionada con la energía cinética de sus moléculas o rapidez del movimiento de las partículas que constituyen la materia, a mayor agitación, mayor será la temperatura. Un instrumento que se utiliza para medir la temperatura se llama termómetro. La temperatura ambiente promedio de las mediciones realizadas en un lapso de 10 minutos (se toman muestras de cada minuto), se mide en unidades dependientes de la escala elegida, que pueden ser: Celsius ( $^\circ\text{C}$ ) es la más común, Fahrenheit ( $^\circ\text{F}$ ) y Escala Kelvin ( $^\circ\text{K}$ ) (Oldani, 2020) (Fundación Española para la Ciencia y la Tecnología, 2004).

La humedad es la cantidad de vapor de agua en el aire. El contenido de humedad de la atmósfera se expresa de varias formas, tales como: humedad absoluta, humedad específica y la relación de mezcla. Sin embargo, el método más común para medir la humedad es la llamada humedad relativa, que se expresa en porcentaje (%). La humedad generalmente se mide con un instrumento llamado psicrómetro o higrómetro y se calcula a partir de la expresión:

$$h = \frac{e}{E} 100$$

En esta expresión **e** representa el contenido de vapor de la masa de aire y **E** su máxima capacidad de almacenamiento de éste, llamada presión de vapor saturante (Fundación Española para la Ciencia y la Tecnología, 2004).

La presión atmosférica es la fuerza que el aire que nos rodea ejerce sobre todos los objetivos a la gravedad por unidad de área, depende de muchas variables, incluida la altitud. Un instrumento utilizado para medir la presión atmosférica es un barómetro. Viene en una variedad de unidades, siendo la más comunes: presión atmósferas, milímetro(mm) de mercurio, pascales, hectopascales y milibares. Se puede considerar una conversión entre los dos: 1 atmósfera = 760 mmHg = 101300 N/m<sup>2</sup> (o Pa) = 1013 mb (o hPa). En una Estación Meteorológica Automática se obtiene del promedio de las mediciones realizadas en un lapso de 10 minutos (se toman muestras cada minuto), y se mide en hPa (Hectopascales) (Fundación Española para la Ciencia y la Tecnología, 2004).

La radiación solar es la energía emitida por el Sol hacia la Tierra que viaja por el espacio en forma de ondas electromagnético. Las ondas con mayor energía son las del rango ultravioleta, seguidas de la luz visible, infrarrojos, etc. Así hasta las menos energéticas que corresponden a las ondas de radio. La cantidad de radiación solar recibida en un punto se mide mediante un aparato llamado piranómetro, que registra la duración y la intensidad se usa el heliógrafo. Su unidad es el vatio por metro cuadrado y la medición es el W/m<sup>2</sup> (watt/m<sup>2</sup>) (Fundación Española para la Ciencia y la Tecnología, 2004).

A modo de resumen en la siguiente Tabla 4, se muestra un resumen de los instrumentos de medida y su correspondiente variable meteorológica a medir y las unidades de medida respectiva.

**Tabla 4***Instrumentos de variables meteorológicas*

<b>Instrumento</b>	<b>Variable meteorológica</b>	<b>Unidad de medida</b>
Termómetro	Temperatura	°C
Barómetro	Presión atmosférica	hPa
Pluviómetro	Precipitación	l/m <sup>2</sup>
Higrómetro	Humedad relativa	%
Evaporímetro	Evaporación	Mm de agua evaporada
Anemómetro	Velocidad del viento	m/s – km/h
Veleta	Dirección del viento	°
Heliógrafo	Horas de sol	h
Radiómetro	Radiación	w/m <sup>2</sup>

**b. Sistemas de observación de variables meteorológicas**

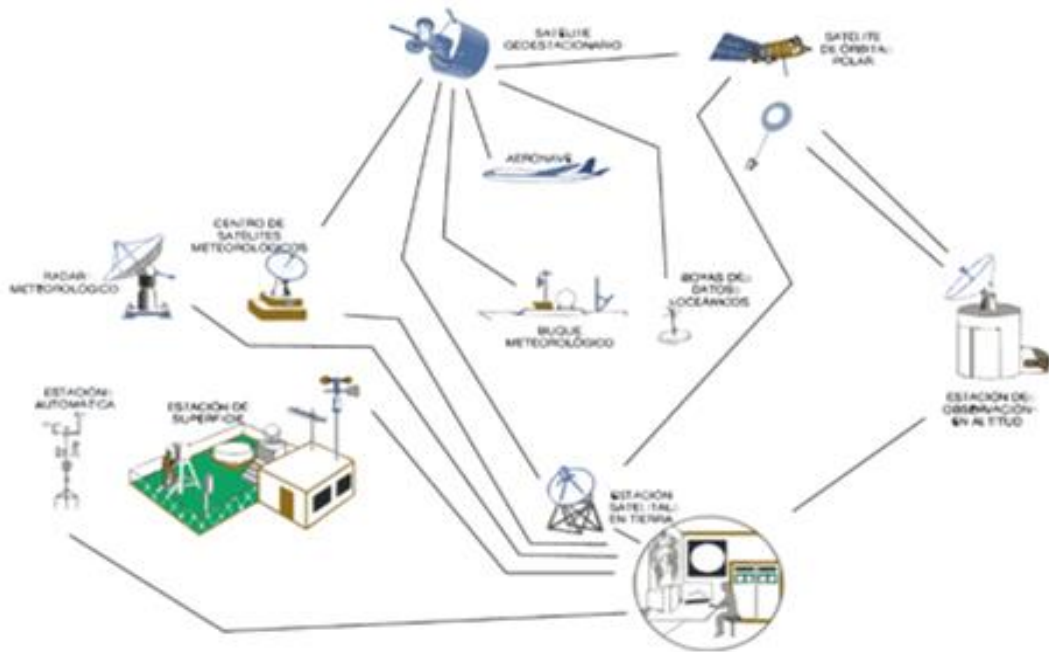
Para el estudio y medición de variables meteorológicas existen diversos sistemas de observación de medición de datos basados en las consideraciones temporales y espaciales necesarias. Utilizando estos sistemas e instrumentos de observación, se pueden obtener indicadores de estas variables desde diferentes perspectivas y con diferentes errores, permitiendo una mejor comprensión del comportamiento de las variables meteorológicas a escala global. Los satélites meteorológicos son el sistema de observación más complejo porque están en órbita alrededor de la Tierra, lo que les otorga un campo de visión privilegiada, más amplia y completo que cualquier aparato situado en la superficie terrestre (Fundación Española para la Ciencia y la Tecnología, 2004).

Uno de los sistemas de recolección de datos más utilizados es la Estaciones Meteorológicas Automáticas – EMA (*Automatic Weather Station - AWS*), que es un sistema basado en radar o detecciones desde satélites donde las observaciones y transmisiones se realizan de manera automática, como se muestra en la Figura 1. Las variables meteorológicas medidas mediante los sensores que se encuentran integrados son leídas o recibidas por una unidad central de adquisición de datos, y un grupo de dispositivos eléctricos que realizan mediciones de las variables meteorológicas de manera automática se denomina Estación Meteorológica

Sinóptica (ESIME), como se muestra en la Figura 4 (Mendoza et al., 2020; Guerra, 2017).

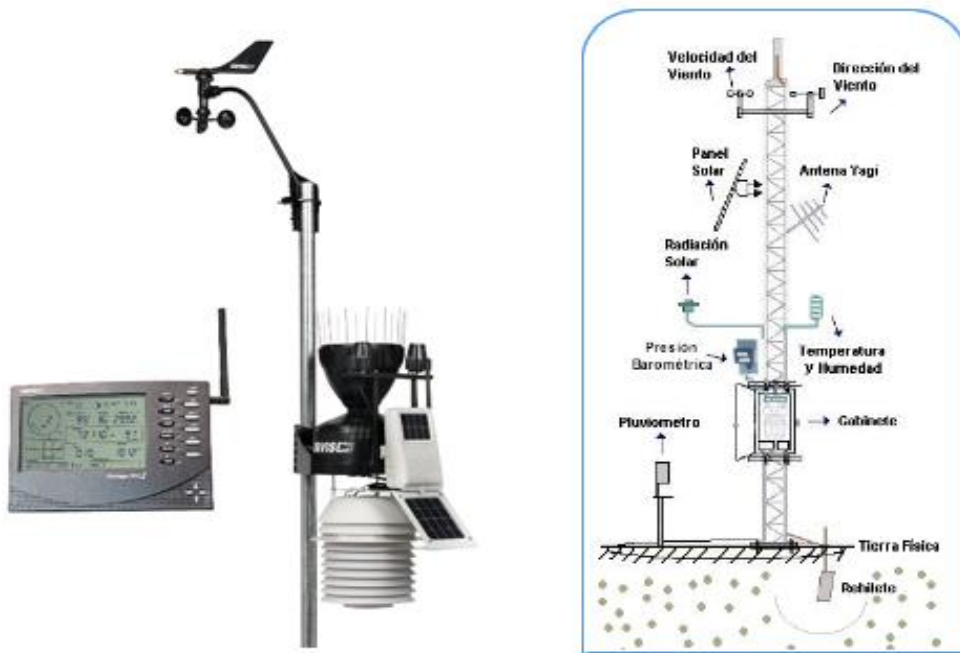
**Figura 4**

*Sistema de observación meteorológica*



**Figura 5**

*Estación Sinóptica Meteorológica (ESIME)*



### **c. Patrones de comportamiento meteorológico**

El comportamiento en la naturaleza se basa en ciclos, tenemos ciclo de agua o ciclo de vida, también se conocen ciclos climáticos, las glaciaciones se alternan con interglaciares cálidos, en meteorología pasan cosas parecidas, hay patrones de comportamiento. La meteorología estudia la atmósfera desde unas perspectivas diferente. Por un lado, describe las condiciones del tiempo atmosférico en un espacio y tiempo concreto. Por otro lado, se desarrolla investigaciones sobre el comportamiento de cada uno de las variables meteorológicas (temperatura, presión, humedad, radiación solar, etc.) para descubrir las leyes y patrones de su influencia sobre otras variables y realizar una previsión del tiempo acertada.

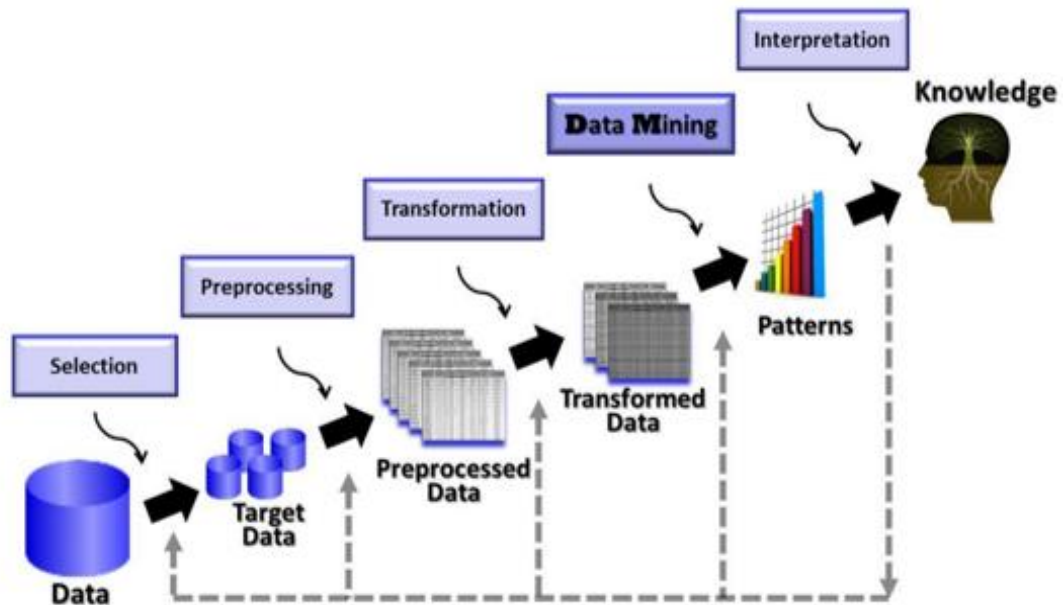
Los fenómenos y cambios meteorológicos influyen diariamente a la sociedad y al medio ambiente con efecto negativo como beneficiosos por lo tanto, la importancia de estudiar datos meteorológicos para descubrir o comprender su comportamiento, con el objetivo de caracterizar en determinados periodos la frecuencia de ocurrencia de los fenómenos, clasificarlos, observar comportamientos, descubrir patrones y realizar estimaciones si fuera posible, como por ejemplo “cielos despejados”, “humo”, “brumas”, “neblinas”, “nieblas”, “relámpagos”, “lluvias”, “chubascos” y “tormentas eléctricas” para periodos estacionarios o periodos en un determinado día (Alvarez & Borrajero, 2016).

### **d. Técnicas para la determinación de patrones de comportamiento.**

Es frecuente usar el termino KDD (Knowledge Discovery in Databases) como sinónimo de Data Mining o minería de datos, pero existen diferencias muy claras entra ambos, KDD es un proceso que consta de una serie de fases, mientras que Data Mining es tan sólo una de estas fases, pero es considerado como el núcleo del proceso KDD tal como se muestra en la Figura 6 (Zamora, 2018).

**Figura 6**

*Etapas KDD. Data Mining es el núcleo de un proceso de descubrimiento de conocimiento*



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

### **e. Metodologías, Técnicas y aplicaciones de Data Mining a datos meteorológicos.**

Data Mining es el proceso de extraer conocimiento valiosa, útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos (Mulatu et al., 2020). Esta técnica tiene un enfoque orientado no solo a la recopilación de datos, incluye también el procesamiento previo de datos y descubrimiento de conocimientos (Knowledge Discovery Approach KDD), es muy reciente la aplicación de esta técnica en el campo del análisis de datos meteorológico (Big data) recopilados a través de sistemas de observación e instrumentos de medición meteorológica, como estudio científico de la atmósfera (Khan et al., 2016).

Muchas técnicas de Data Mining para pronosticar el clima fueron utilizados hasta hoy tales como las técnicas de clasificación y de agrupamiento, sin embargo, las redes neuronales artificiales (Artificial Neural Network - ANN) resultaron ser la mejor opción para la predicción del tiempo, ya que esta técnica permiten capturar

fácilmente relaciones complejas en sistemas lineales y no lineales sin la necesidad de suposiciones como se necesitaba en los enfoques tradicionales (Yusf, 2018).

Se utilizaron técnicas de minería de datos en la Cujae – Cuba con el objetivo de obtener modelos de datos utilizando el método CRISP-DM y herramienta de análisis de datos WEKA para estimar el comportamiento de parámetros meteorológicos secundarios en datos de superficie. Entre ellos, la selección de atributos y las tareas de regresión utilizan tecnología de red neuronal perceptrón multicapa (Cruz, Zerquera, Morales, & Rosete, 2012).

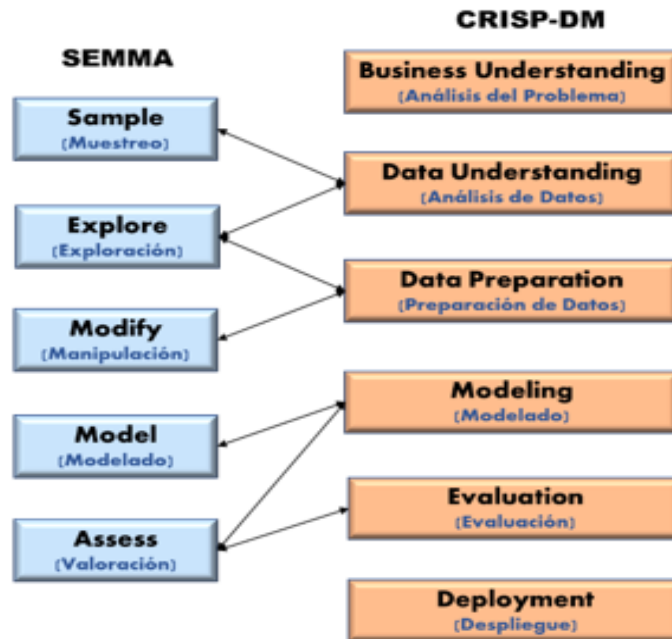
Se utilizaron herramientas computacionales de minería de datos de Watson Analytics para identificar y extraer patrones de comportamiento o generativos en datos climatológicos, como la precipitación registrada por el Servicio Meteorológico Nacional (SMN) en los estados de Chiapas, Oaxaca, Tabasco y Veracruz. Método de minería de datos Watson Analytics se utilizó la metodología de Data Mining CRISP-DM, con el objetivo de tomar medidas y formular estrategias de prevención ante cualquier eventualidad ocasionada por el clima (Castorena, 2018).

Otra enfoque de proyectos para procesamiento de datos mediante la minería de datos es el enfoque Data Mining es la metodología SEMMA ((Sample, Explore, Modify, Model, Assess) que al igual que la metodología CRISP-DM comparten la misma esencia, la primera se centra más en las características técnicas del desarrollo del proceso, que el SAS Institute lo ha desarrollado, mientras que CRISP-DM, mantiene una visión más amplia respecto a los objetivos del proyecto de minería de datos, esta última viene implementado en el IBM SPSS Modeler, en la Figura 7, se muestra las fases e interrelaciones de éstas dos metodologías (Montequín, 2002; Zamora, 2018).



**Figura 7**

*Esquema comparativo de las tareas generales y fases de las Metodologías de Data Mining SEMMA y CRIP-DM*



**Nota.** Adaptado de Montequín (2002).

#### **f. Data Mining Algoritmos de aprendizaje supervisado y no supervisado**

Data Mining para lograr sus propósitos utilizan una gran cantidad de algoritmos, organizados según el tipo de aprendizaje, utilizado que pueden ser tanto supervisados como no supervisado. Los algoritmos de aprendizaje supervisado aprenden de un conjunto de datos de entrenamiento previamente etiquetados con variables donde se sabe que el valor de un atributo de destino predice o la variable dependiente que asocian con la entrada y la salida; nuevamente, estos algoritmos pueden ser algoritmo de regresión o de clasificación. Los algoritmos de aprendizaje no supervisado basan su proceso de entrenamiento en un conjunto de datos sin etiquetas, variables o clases previamente definidas, su función es agrupar, segmentación (clúster) para encontrar grupos similares en el conjunto de datos, estos grupos pueden ser jerárquicos o particionales (García, 2020).

### **g. Análisis de series temporales**

Una serie temporal, serie temporal o cronológica es una secuencia de datos medidos a intervalos de tiempo sucesivos y espaciados, usualmente de manera uniforme, donde  $Z(t)$  es un conjunto de observaciones secuencialmente realizadas en el tiempo, de modo que le corresponde un valor  $Z_t$  a cada instante de tiempo  $t$  observado (Tak-chung, 2011). Usando el lenguaje matemático, una serie de tiempo es considerada como una colección de variables aleatorias  $\{Z_t, t \in T\}$  donde  $T$  es un conjunto de índices, normalmente el conjunto de los números naturales, por tanto la serie de tiempo de  $n$  observaciones sucesivas será  $(Z_1(t), Z_2(t), \dots, Z_n(t))$  los valores de la serie pueden ser vistos como salidas de un proceso aleatorio, esto significa que cada valor  $Z_t$  de la serie de tiempo puede ser considerado como una observación de una de las variables aleatorias  $Z_t$  que integran el proceso (Hervás, 2020; Soto, 2013).

En un análisis de series temporales se estudian fenómenos secuencialmente ordenados en el tiempo, en la que se tiene en cuenta la dinámica de los procesos con la finalidad de entenderlos de la mejor manera posible, los datos en una serie de tiempo tienen un orden natural, esto hace que su análisis sea un tanto distinto al de otros problemas que no presentan un orden natural en sus observaciones (Soto, 2013).

### **h. Clasificación de series temporales en WEKA**

El software WEKA (Waikato Environment for Knowledge Analysis|) es una aplicación de código abierto escrito en Java, creado en la Universidad de Waikato en Nueva Zelanda, y publicado bajo la Licencia Pública General de GNU, cuenta con algoritmos de aprendizaje automático para tareas de análisis de datos y modelado predictivo y además contiene una interfaz gráfica de usuario que facilita el acceso a sus funciones de análisis de datos (Mulatu et al., 2020; Soto, 2013).

En Weka Software se encuentra disponible un paquete denominado **timeseriesForecasting** el cual está dedicado específicamente a la predicción de series temporales, este paquete utiliza una máquina de aprendizaje automatizado para crear los modelos de la serie de tiempo y las predicciones del

comportamiento futuro de la serie, permitiendo el desarrollo de modelos predictivos de los datos, y además su evaluación y visualización de forma clara y precisa. A pesar de los alcances por Weka software como un poderoso sistema de aprendizaje automatizado, según Soto (2013) en su investigación concluye que Weka ofrece muy pocas posibilidades para el trabajo con series temporales.

#### **i. Análisis de series temporales con RStudio software**

“R Project for Statistical Computing” es una herramienta focalizada al análisis estadístico de datos y permite realizar gráficas, operaciones matemáticas y cálculos en general, posee un entorno y lenguaje de programación, ya que, “R” permite dar instrucciones mediante códigos para que el software realice diversas tareas tales como: realizar operaciones con multitud de tipos de datos, hacer cálculos y realizar gráficos para su interpretación. “RStudio” es un entorno integrado de desarrollo (Integrated Development Environment - IDE) que ofrece una forma más cómoda e intuitiva de trabajar con R (Hervás, 2020).

En análisis de series de tiempo es de vital importancia en diferentes campos de la investigación a la hora de elaborar predicciones, un estudio realizado por Hervás (2020) en el que focaliza su análisis en el uso de la metodología de Box & Jenkins a través del software estadístico “RStudio”, ya que, a pesar de ser un software gratuito, se trata de una de las herramientas más importantes y potentes a la hora de realizar análisis estadísticos y en la que se revisan las principales funciones disponibles para el análisis de series temporales.

#### **j. Data Mining para Series Temporales**

Los modelos basados en Data Mining para series temporales son contribuciones importantes para el análisis de datos, en este caso para datos meteorológicos ya que las variables de estudio son obtenidas en series de tiempo periódicas, debido a que son capaces de caracterizar satisfactoriamente series periódicas, no periódicas, complejas y caóticas. Estos métodos cubren las limitaciones de las técnicas tradicionales utilizadas en el análisis de series temporales, ya que adaptan los conceptos de Data Mining para tratar series como una clase especial de datos. Su campo de estudio utiliza lo mejor de las siguientes áreas de investigación: análisis estadístico de series temporales, Data Mining, procesado

adaptativo de señales, análisis ondulatorio, algoritmos genéticos, sistemas dinámicos y caos (Soto, 2013).

#### **k. Análisis de datos y modelos de predicción mediante IBM SPSS Modeler software**

SPSS Modeler Subscription Software es desarrollado por IBM Corp. Originalmente se llamaba SPSS Clementine y es un programa especializado en el análisis de grandes bases de datos de diferente tipo orientado a desarrollar procesos de Data Mining con el objetivo de comprender mejor a los datos y obtener los mejores resultados mediante una interfaz visual que conlleva a modelos predictivos muy útiles en el análisis y procesamiento de datos meteorológicos más eficaces y en menor tiempo (pronósticos, clasificación, segmentación y asociaciones).

SPSS Modeler en su diseño considera la metodología CRISP-DM y para el desarrollo de proyectos de Data Mining considera una interfaz gráfica muy amigable y las diferentes funciones están basados en nodos, iconos que se van tomando y formando una ruta o stream que puede archivar individualmente o por proyectos. La ruta de datos consiste en una secuencia de operaciones donde los datos fluyen de registro en registro, se manipulan hasta llegar a un destino y como resultado final se obtiene un nugget que contiene todas las operaciones realizadas (Mejia, 2019).

#### **l. Técnica del Árbol de decisión y su aplicación a datos meteorológicos.**

El árbol de clasificación o árbol de decisión es una técnica de Data Mining supervisada que se puede utilizar en muchas áreas de la ciencia, por ejemplo esta técnica se utilizó en la provincia de Chimborazo – Ecuador con el objetivo de identificar variables meteorológicas predominantes a ocho intervalos de tiempo, utilizando el software estadístico R con datos de 2016 y publicado en 2020, las variables analizadas fueron: temperatura del aire, humedad relativa, presión barométrica, radiación solar difusa, radiación solar global, temperatura del suelo a -20cm. y velocidad del viento. El árbol de decisión mostró que la principal variable en esta zona es la radiación solar global y se puede predecir la

temperatura del aire con una probabilidad del 66%, y se puede clasificar los datos con un 71% de probabilidad (Haro, 2020).

### 2.3 Marco filosófico

#### **Encuadramiento Epistemológico del Problema**

Luego de proponer los objetivos generales y específicos del problema de investigación y su solución, se enmarca epistemológicamente la solución del problema tratando de identificar el tipo de normas o tendencia en la que se sustenta a partir de las características del problema. Explicar y tratar diferentes direcciones filosóficas, de manera que las preguntas de investigación planteada sean explicables y justificadas.

#### **El Pragmatismo**

El filósofo norteamericano William James (1910) es considerado el fundador del pragmatismo. El pragmatismo rechaza el concepto de verdad en el sentido de que el pensamiento corresponde a la existencia. Pero el pragmatismo no se detiene en esta negación, sino que reemplaza el concepto descartado con un nuevo concepto de la verdad. Según él, lo que realmente significa es útil, valioso y enriquecedor. El pragmatismo modifica así el concepto de verdad porque parte de un cierto concepto de hombre. Según él, el hombre no es ante todo la existencia de la teoría o del pensamiento, sino la existencia de la práctica, la existencia de la voluntad y la acción. El error fundamental del pragmatismo es que no ve el campo lógico, ignora su valor propio y la autonomía del pensamiento humano (Hessen, 2006).

#### **Racionalismo**

La forma más antigua del racionalismo se originó en Platón. El racionalismo es la posición epistemológica de que pensar en la razón es la fuente principal del conocimiento humano, se llama racionalismo (de ratio = razón). Según él, el conocimiento merece realmente ese nombre sólo cuando es lógicamente necesario y universalmente válido. Cuando nuestra razón juzga que algo debe ser así y que no puede ser de otro modo; y por tanto, según el racionalismo, debe ser así siempre, en todas partes, y solo cuando estamos en la verdad ante el conocimiento. El mérito del racionalismo es que ve y enfatiza la importancia del elemento racional en el conocimiento humano. El inconveniente del racionalismo

es que tiene un espíritu del dogmático. Creía que podía entrar en el reino de metafísica a través de líneas de pensamiento puramente conceptuales (Hessen, 2006).

### **El Empirismo**

Su fundador fue John Locke (1632-1704). Lo opuesto al argumento empirista contra el racionalismo (según el cual el pensamiento, la razón, es la verdadera fuente del conocimiento) es este: la única fuente del conocimiento humano es la experiencia. Según el empirismo, no hay ningún patrimonio a priori de la razón. La conciencia cognoscente deriva su contenido no de la razón, sino exclusivamente de la experiencia. El espíritu humano está por naturaleza vacío; es una tabula rasa, una hoja por escribir y en la que escribe la experiencia. Todos nuestros conceptos, incluso los más generales y abstractos, provienen de la experiencia (Hessen, 2006).

### **Intelectualismo:**

El fundador del intelectualismo es Aristóteles. El intelectualismo es una orientación epistemológica que es uno de esos intentos de mediar entre el racionalismo y el empirismo. El racionalismo cree que el pensamiento es la fuente y la base del conocimiento, y el empirismo es la experiencia, mientras que el intelectualismo cree que ambos factores están involucrados en la creación del conocimiento. El intelectualismo está de acuerdo con el racionalismo en que existe juicios lógicamente necesarios y universalmente justificada, y no solo sobre los objetos ideales, (como lo reconoce principales empirismos), sino también sobre los objetos reales. Pero mientras el racionalismo afirma que los elementos de estos juicios, los conceptos, como un patrimonio a priori de nuestra razón, el intelectualismo lo deriva de la experiencia (Hessen, 2006).

### **El Apriorismo**

El fundador de este apriorismo es Kant. El Apriorismo presenta un segundo intento de mediación entre el racionalismo y el empirismo. El Apriorismo también considera la experiencia y el pensamiento como fuentes de conocimiento. Pero el apriorismo define la relación entre la experiencia y el pensamiento de una manera que se opone directamente al intelectualismo. Como

ya dice el nombre de apriorismo, nuestro conocimiento en el sentido de esta dirección presenta elementos a priori independientes de la experiencia. Esta fue también la opinión del racionalismo. Pero, aunque éste consideraba que los factores a priori eran conceptos perfectos, estos factores son formales a priori. No son contenidos, sino formas de información. Estas formas reciben su contenido de la experiencia, y aquí el apriorismo difiere del racionalismo y se aproxima al empirismo. En cierto modo, los factores a priori son como recipientes vacíos que se llena con contenidos concretos. El principio del apriorismo dice: "Los conceptos sin las intuiciones son vacíos, las intuiciones sin los conceptos son ciegas" (Hessen, 2006).

### **El Objetivismo**

Platón fue el primero en defender el objetivismo. Según esta escuela de filosofía, el objeto es decisivo entre los dos miembros de la relación cognitiva. El objeto define al sujeto. Debe ser controlado por él. El sujeto de alguna manera toma las propiedades del objeto y las reproduce. Presupone que el objeto se enfrenta a la conciencia cognoscente como algo completo, algo definido en sí mismo. Esta es exactamente la idea principal del objetivismo. Según él, un objeto es un dato que asume una estructura que se reconstruye, digamos, por medio de la conciencia cognitiva (Hessen, 2006).

### **El Realismo**

El Realismo, la posición epistemológica de que hay cosas reales, independientes de la conciencia. Esta posición permite diferentes formas. primitivo, tanto histórica como psicológicamente, es el realismo ingenuo. Este realismo aún no está afectado por ninguna reflexión crítica acerca del conocimiento. El problema del sujeto y el objeto aún no existe para él. No hace distinción entre la percepción, que es el contenido de la conciencia, y el objeto percibido. No ve que las cosas no nos son dadas en sí mismas, en su corporeidad, inmediatamente, sino sólo como contenidos de la percepción. Y puesto que identifica el contenido de la percepción con los objetos, da a estos últimos todas las propiedades contenidas en los primeros. Según él, las cosas son exactamente como las percibimos. Los colores que vemos en ellos les pertenecen como propiedades objetivas. Lo mismo se aplica a su sabor y olor, su dureza o blandura, etc. (Hessen, 2006).

## **El Fenomenalismo**

La teoría del fenomenalismo fue desarrollada por Kant. El fenomenalismo es la teoría de que no conocemos las cosas como son en sí mismas, sino como se nos aparecen. Para el fenomenalismo, las cosas reales existen, pero no podemos conocer su naturaleza. Solo podemos saber que son las cosas, pero no lo que son. El fenomenalismo coincide con el realismo en el reconocimiento de las cosas reales; pero concuerda con el idealismo limitar el conocimiento a la conciencia, el mundo de las apariencias, lo que conduce inmediatamente a la incognoscibilidad de las cosas en sí mismas (Hessen, 2006).

## **¿Qué es un Metalenguaje?**

Según la Real Academia de Española, metalenguaje se define como una lengua que se utiliza para hablar de una lengua, es decir, es una lengua que se utiliza para hablar de aspectos de una lengua o para describirla como un todo (ASALE, s. f.).

Todo lenguaje tiene la capacidad de tomar como referencia o directriz un objeto. De esta forma, cualquier lengua que se centre en el estudio de otro lenguaje que se considera un metalenguaje. Por lo tanto, se puede decir que las metalenguas se basan en la comprensión adquirida del lenguaje utilizado en cualquier tipo de actividad. Su función es conocer sus componentes para dominarlos. Como señalaron (Tunmer et al., 1984) el metalenguaje es un componente de la metacognición.

## **Sistema Procesador de Operaciones Metalingüísticas (Metalinguistic Operation Processor, o MOP)**

Según Penagos (2005) el marco del Procesador de operaciones Metalingüísticas (MOP) para todos los idiomas tiene la capacidad de extraer declaraciones y definiciones metalingüísticas de documentación técnica utilizando máquinas de estado finitos y algoritmos de aprendizaje automático.

Este sistema difiere de otros sistemas terminográficos en que no se basa únicamente en regularidad sintácticas, morfológicas o semánticas implícitas en la formación de términos, estructuras semánticas o conceptos, sino que utiliza la dimensión más discursiva de los actos de habla donde las expresiones se establecen explícitamente, editados o calificados por los propios participantes de



la actividad científica. Una vez seleccionadas las expresiones, el sistema realiza un procesamiento lingüístico para crear bases de datos terminológicas semiestructuras denominadas bases de datos de Información Metalingüística (MID).

## CAPÍTULO III

### RESULTADOS, DISCUSIONES Y CONCLUSIONES

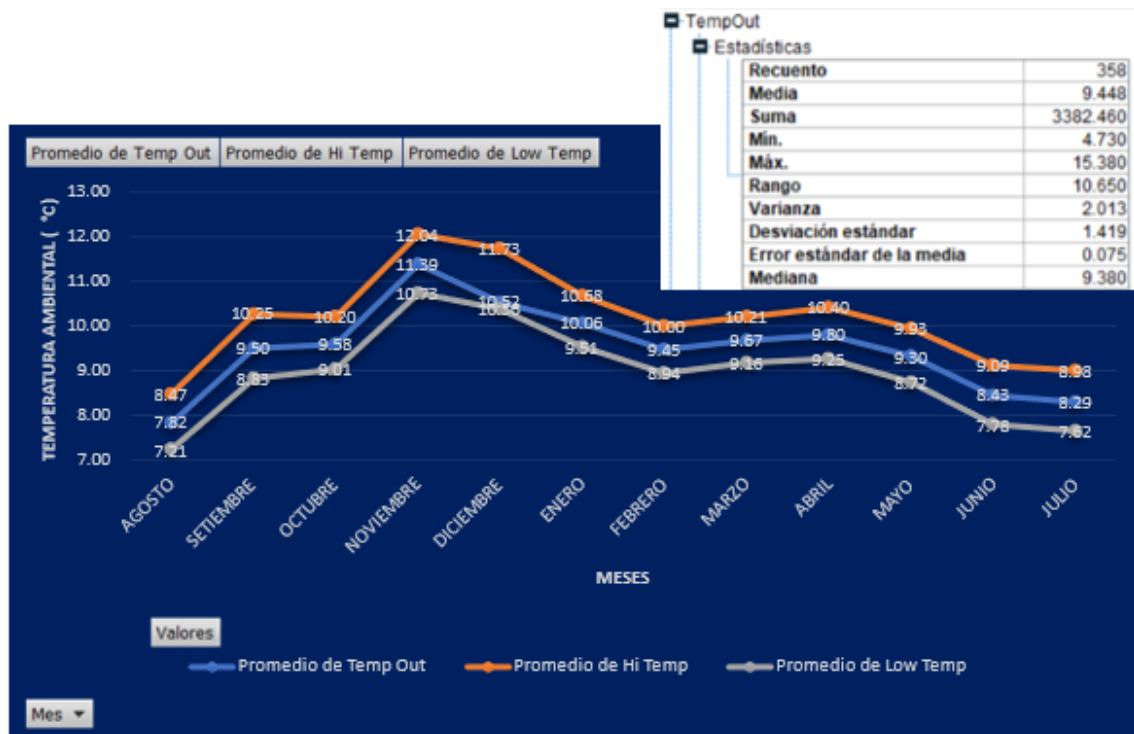
#### 3.1. Exposición resultados

##### 3.1.1. Resultado 1

Los resultados obtenidos del comportamiento meteorológico, asignado con la variable “Temp Out” en la estación meteorológica automática, la misma que muestra que sigue un comportamiento esperado para la ciudad de Huancavelica, los menores registros corresponden a los meses de invierno y a partir del mes de agosto se observa un incremento el que corresponde a los meses de inicio de primavera, observándose también la temperatura promedio más alta en el mes de noviembre. Cómo se observa en la Figura 8, algunos patrones irregulares se muestran en la misma, como por ejemplo los meses con temperaturas mayores en aquellos meses donde se espera que las temperaturas medias sean menores.

**Figura 8**

*Temperatura Ambiental de la ciudad de Huancavelica, periodo 2018-2019.*

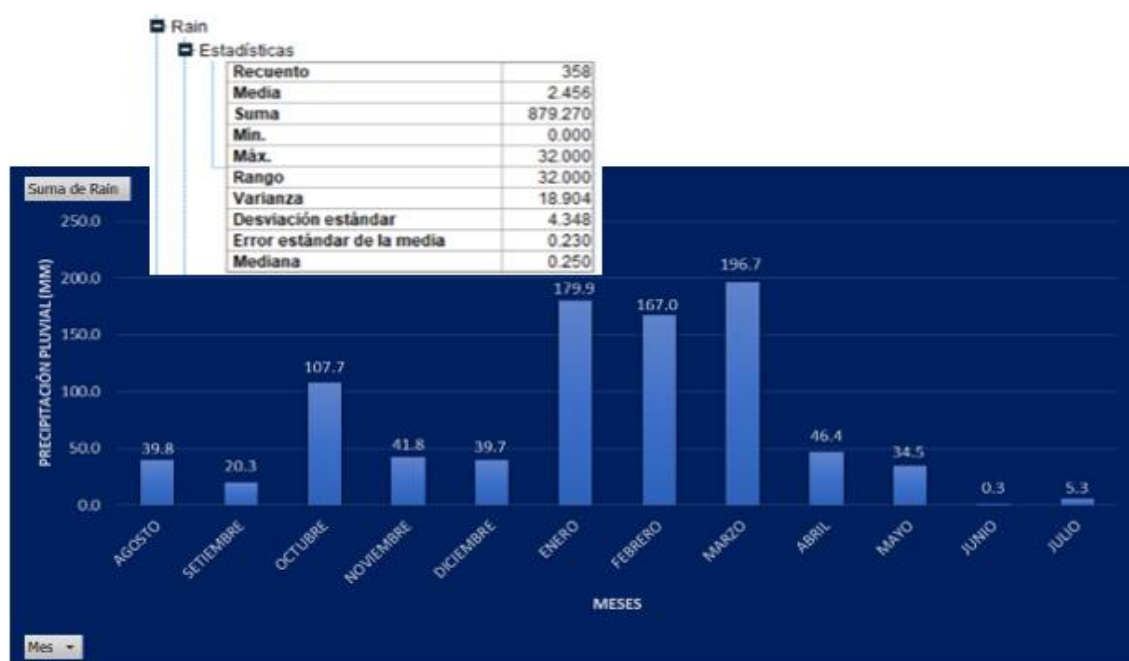


### 3.1.2. Resultado 2

La precipitación pluvial calculado de las observaciones de cada hora del día de estos y el acumulado durante cada mes, asignado con la variable “Rain” en la estación meteorológica automática las mismas, presentan los mayores valores de precipitación pluvial acumulada para la época de verano correspondientes a los meses de enero, febrero y marzo, y los menores registros se presentan durante los meses de invierno. Así también se aprecia la irregularidad de su comportamiento en el periodo de estudio como una evidencia y alerta de lo que se produce en el acumulado de lluvias tal como se observa en la figura 9.

**Figura 9**

*Precipitación Pluvial de la ciudad de Huancavelica, periodo 2018-2019.*

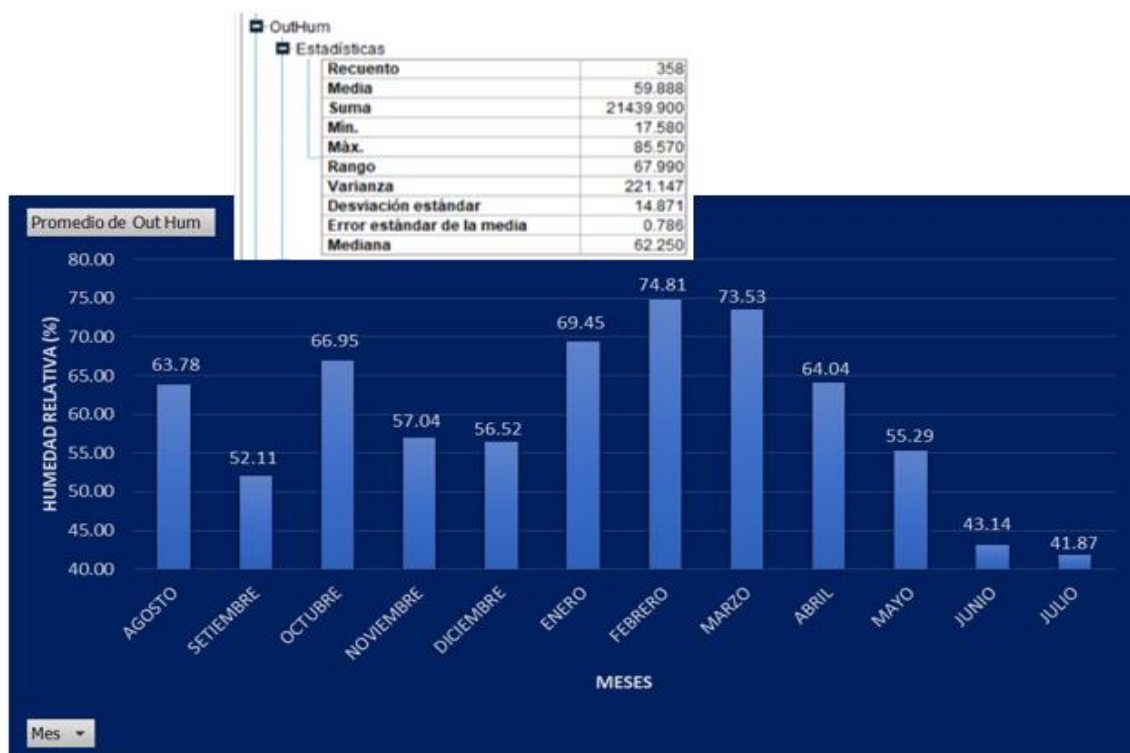


### 3.1.3. Resultado 3

Las humedades relativas medias porcentuales mensuales, asignado con la variable “Out Hum” en la estación meteorológica automática, las mismas que corresponden los valores más altos en los meses de verano (enero, febrero y marzo) tal como se muestra en la Figura 10, que siguen un patrón esperado de los últimos años, descendiendo de manera no tan considerable en las otras estaciones del año, llegando a un promedio mínimo de 41.87 % de humedad relativa en el mes de julio.

**Figura 10**

*Humedad relativa de la ciudad de Huancavelica, periodo 2018-2019.*



#### 3.1.4. Resultado 4

La presión atmosférica media mensual asignado con la variable “Bar” obtenidos en la estación meteorológica automática en milibares y que para su representación gráfica fue convertida por su equivalente en milímetros de mercurio (mm Hg), el comportamiento de esta variable muestra un descenso a partir de septiembre hasta abril, posteriormente se presenta un cambio de comportamiento a partir de mayo cuando termina la estación de verano tal como se muestra en la Figura 11. En el caso de la presión atmosférica horaria se observa que entre las 3pm y 4pm se observa las presiones más bajas en promedio durante un día en periodo de estudio como se muestra en la Figura 12.

**Figura 11**

*Presión atmosférica mensual de la ciudad de Huancavelica, periodo 2018-2019.*



**Figura 12**

*Variación horaria de la Presión atmosférica en la ciudad de Huancavelica, periodo 2018-2019.*

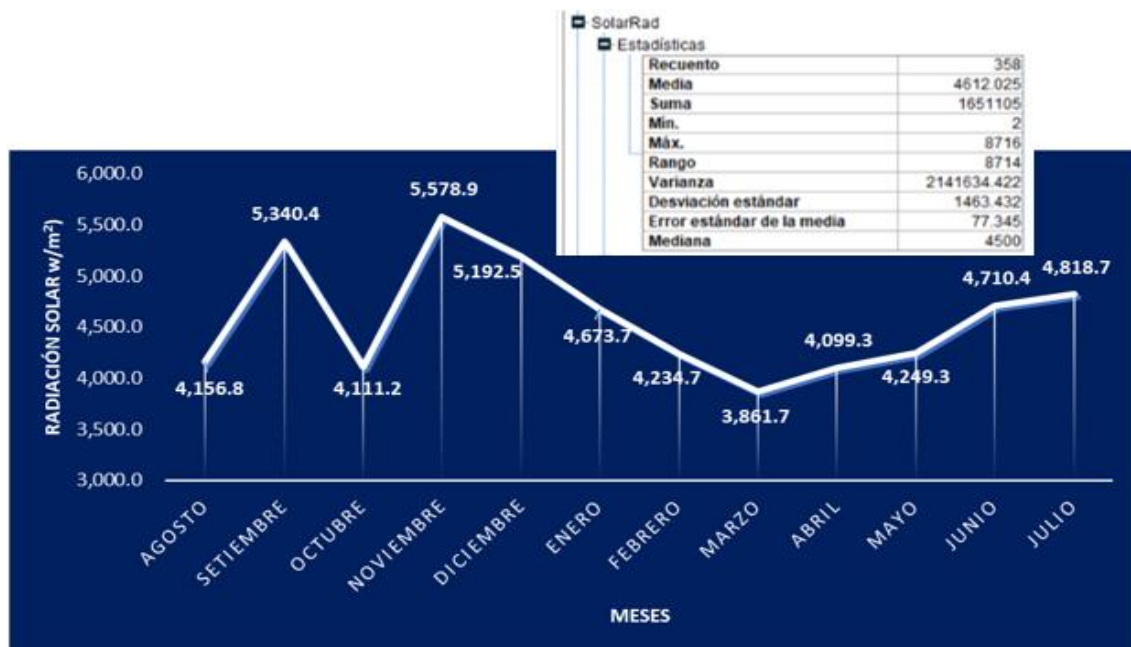


### 3.1.5. Resultado 5

La radiación solar total del día, se consideró los registros acumulados en cada hora durante el día, asignado con la variable “Solar Rad” obtenidos en la estación meteorológica automática en vatios por metro cuadrado ( $W/m^2$ ) mostrándose que el comportamiento es también el esperado, observándose que los registros más bajos, se presentan en la época de verano tal como se muestra en la Figura 13, en el cual presenta alta nubosidad predominante de esta época de lluvia por el cual la radiación es menor que en los otros meses del periodo de estudio, en la Figura 14, se muestra el comportamiento de la radiación solar promedio por cada hora del día durante el año en vatios por metro cuadrado ( $W/m^2$ ), así como también la radiación acumulada de los promedios por cada hora del día durante el año de estudio en megajulios por metro cuadrado ( $Mj/m^2$ ), mostrándose que la radiación máxima durante el día se presenta a las 12 del mediodía y la radiación acumulada promedio por día durante el periodo de estudio fue de  $16.44 \text{ MJ}/m^2$ .

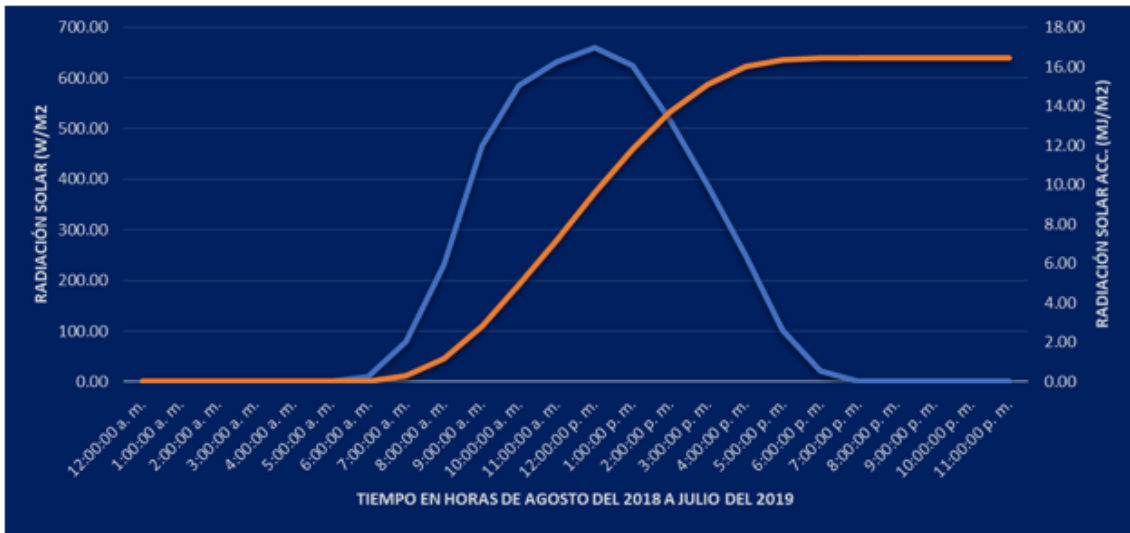
**Figura 13**

*Radiación Solar de la ciudad de Huancavelica por meses, periodo 2018-2019.*



**Figura 14**

*Radiación Solar de la ciudad de Huancavelica por horas, periodo 2018-2019.*



3.1.5. Resultado 5

El índice de radiación ultravioleta, asignado con la variable “UV Index”, el mayor valor encontrado en el día durante en los diferentes meses del periodo de estudio para desarrollar el gráfico, los valores encontrados mantienen una tendencia respecto a un índice de 16 y menores valores ante la presencia de nubosidad en los meses de junio y julio como se muestra en la Figura 15, en la Figura 16 se muestra el comportamiento del índice de radiación ultravioleta promedio en diferentes horas durante el día, mostrándose que los valores máximos se encuentran entre las 10 a.m. al 3 p.m.

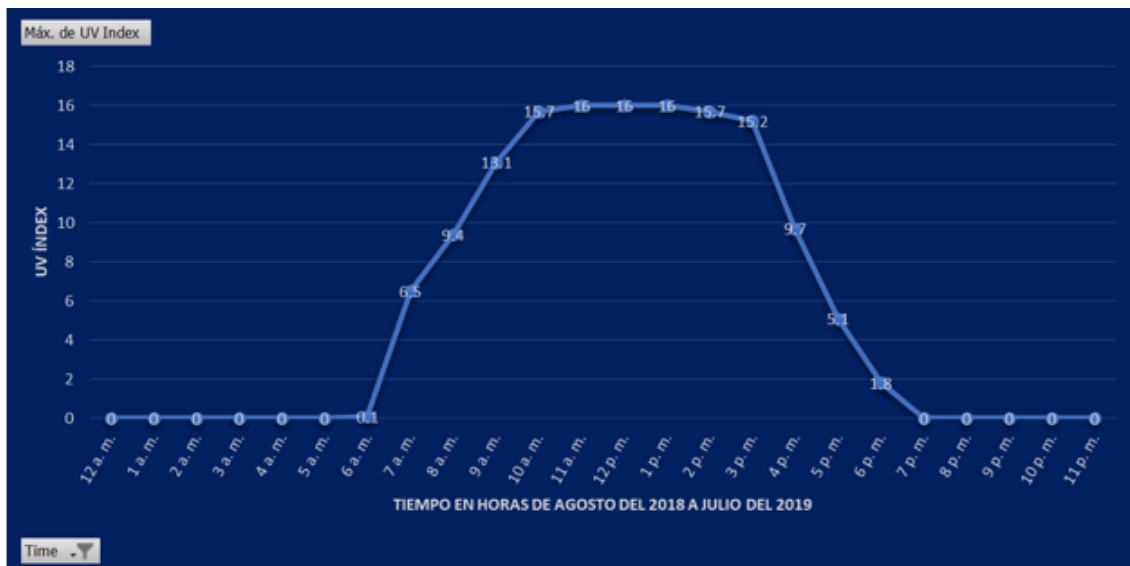
**Figura 15**

*Radiación Ultravioleta de la ciudad de Huancavelica por meses, periodo 2018-2019.*



**Figura 16**

*Radiación Ultravioleta de la ciudad de Huancavelica por horas, periodo 2018-2019.*



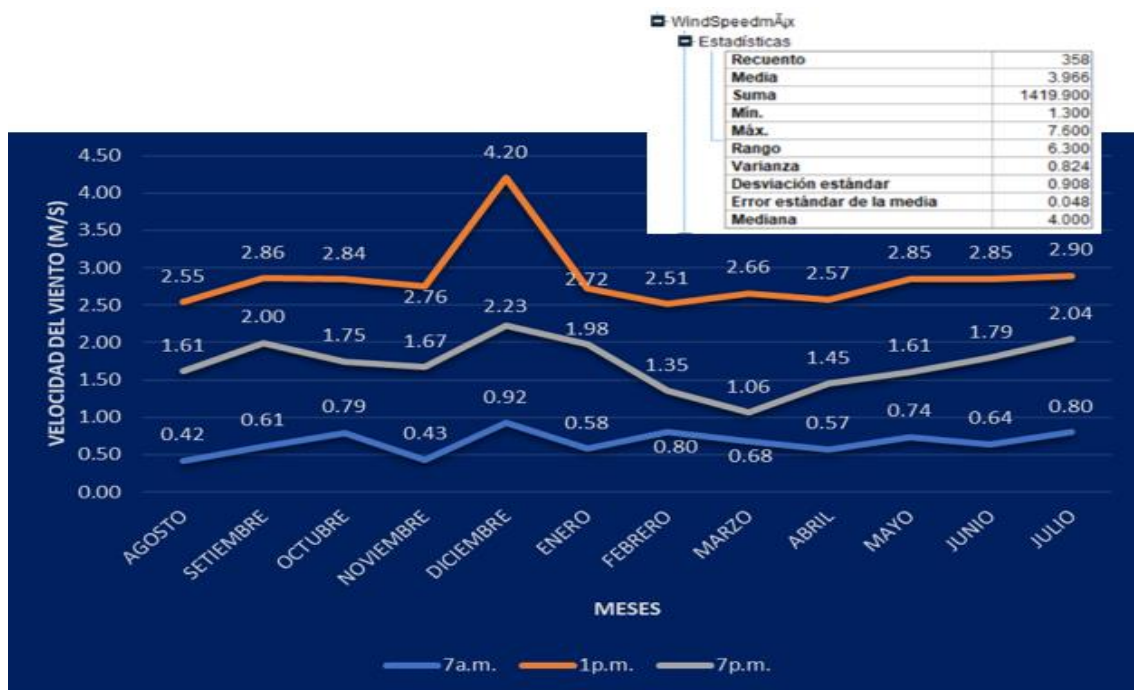


### 3.1.6. Resultado 6

La velocidad del viento, asignado con la variable “Wind Speed”, los datos de los registros considerados en la Figura 17 corresponden a los valores promedios de la velocidad del viento durante un mes en las siguientes horas del día: 7, 13 y 19 horas, los mayores registros corresponden a los meses de la época de invierno y durante el día las velocidades promedio mayores corresponden a las 13 horas y los menores a las 7 horas.

**Figura 17**

*Velocidad promedio del viento mensual a las 7, 13 y 19 horas en la ciudad de Huancavelica, periodo 2018-2019.*

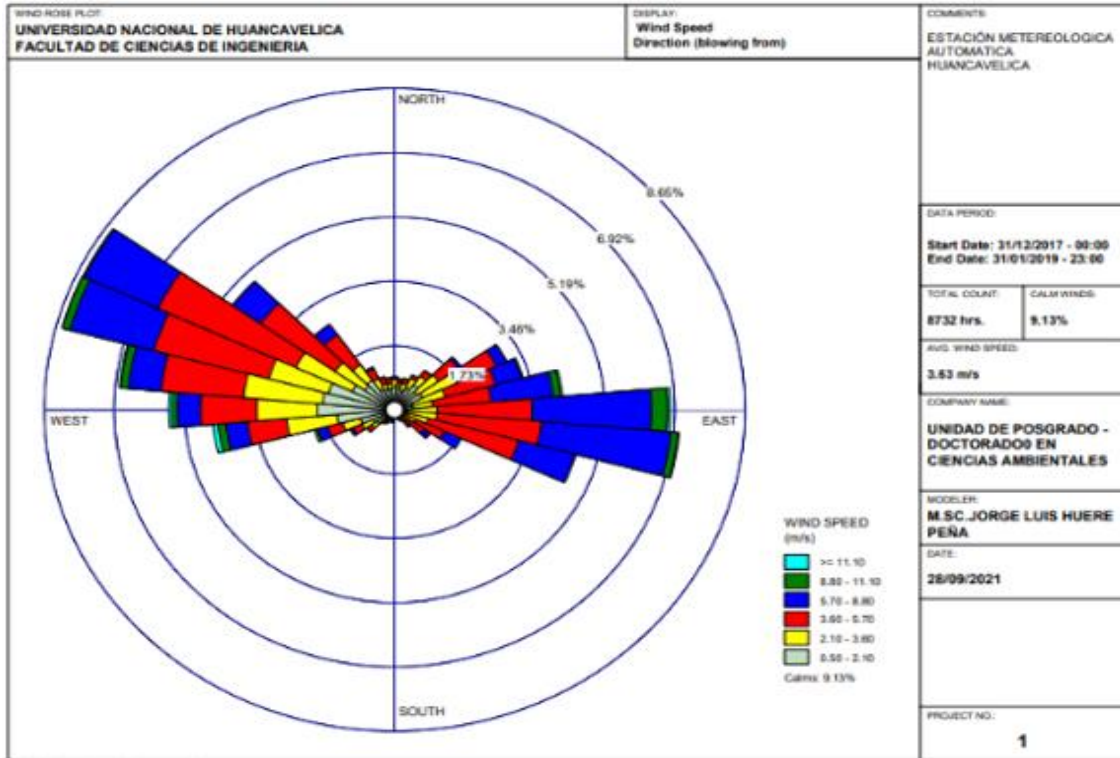


### 3.1.7. Resultado 7

La figura 18 muestran Rosa de viento se puede apreciar el origen del viento (blowing from), y la figura 19 muestran Rosa de viento se puede apreciar la dirección a dónde va el viento (blowing to) la Rosa de Viento (Wind Rose) fueron elaborados a partir de datos registrados en la estación meteorológica automática de 358 días del año que duró el periodo de estudio; en donde se observó que el patrón predominante viene a ser el NO (Nor Oeste) siendo este el origen del soplo de viento. El valor de velocidad de viento se encuentra en el rango de 2.51 m/s hasta 4.20 m/s, teniendo los valores máximos a las 13 horas durante el mes de diciembre.

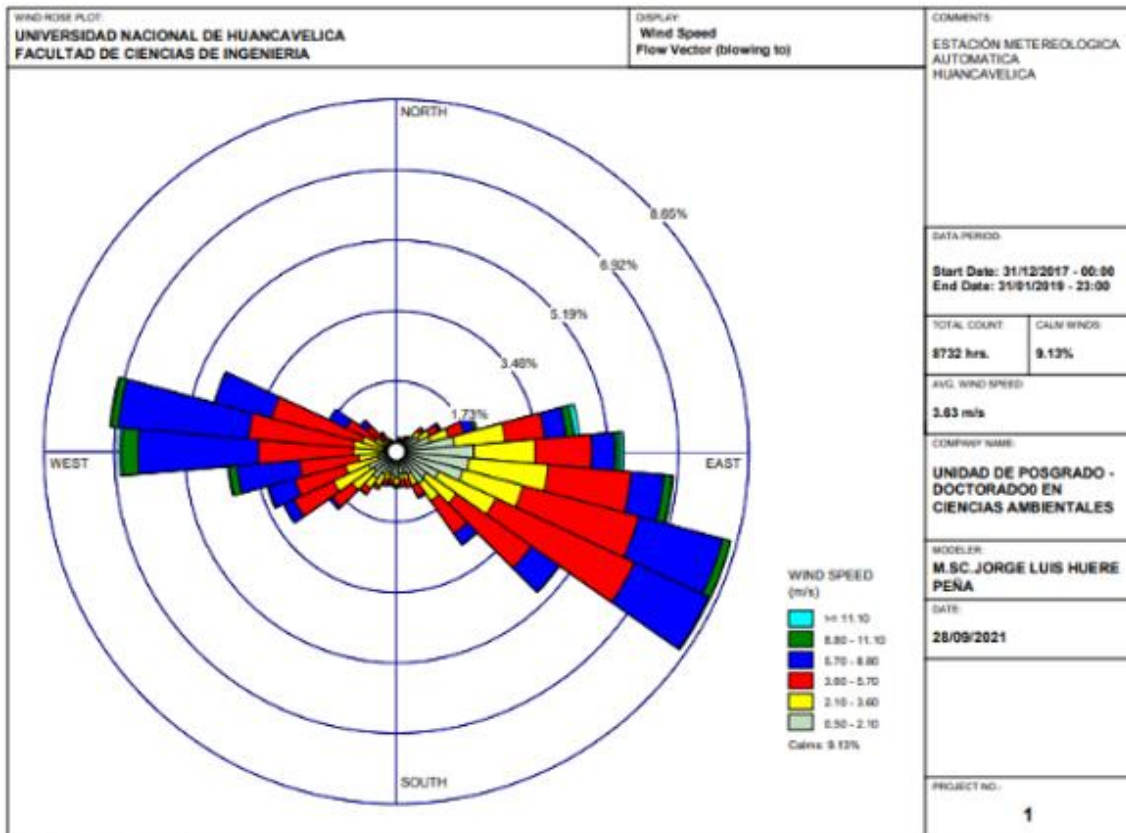
**Figura 18**

Rosa de viento: “blowing from” en la ciudad de Huancavelica- periodo 2018-2019.



**Figura 19**

*Rosa de viento: “blowing to” en la ciudad de Huancavelica- periodo 2018-2019.*

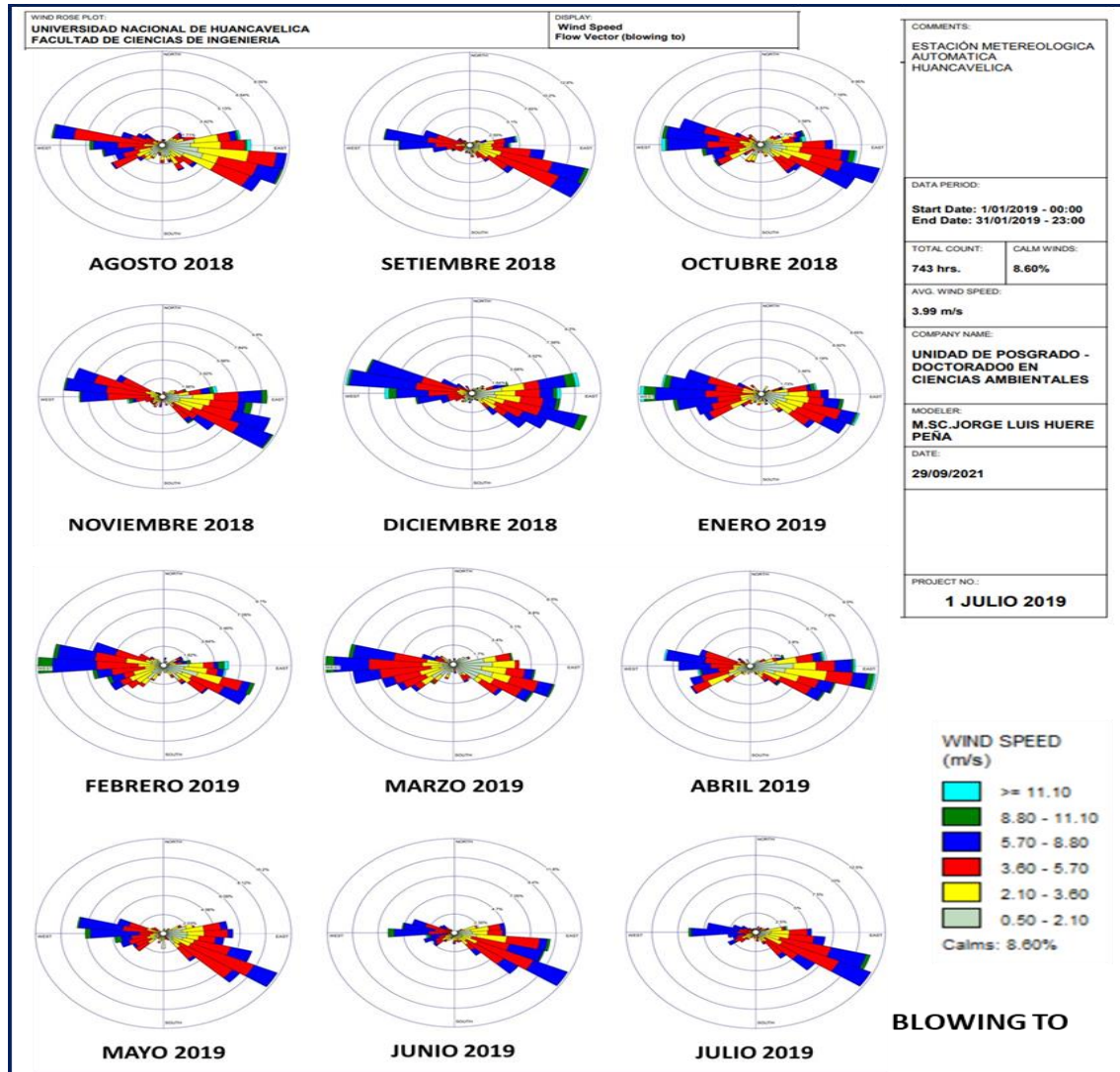


### 3.1.8. Resultado 8

La Figura 20 muestran las Rosas de vientos (Wind Rose) en las que se puede apreciar la dirección de donde proviene el viento (blowing from) y sus variaciones durante los 12 meses que duró el estudio y la Figura 21 muestra la dirección a dónde va el viento (blowing to), también durante un periodo de 12 meses, se observó que el patrón predominante fue el NO (Nor Oeste) en el caso de blowing from y el patrón predominante en el caso de blowing to fue el SE (Sur Este). estas imágenes fueron obtenidas mediante WRPLOT de Lakes Environmental Software que requiere de la base de datos las variables Fecha, Velocidad del viento, Dirección del viento y Precipitación pluvial para el diseño de la rosa de viento.

**Figura 20**

Rosa de viento: “blowing from” en la ciudad de Huancavelica por meses 2018-2019

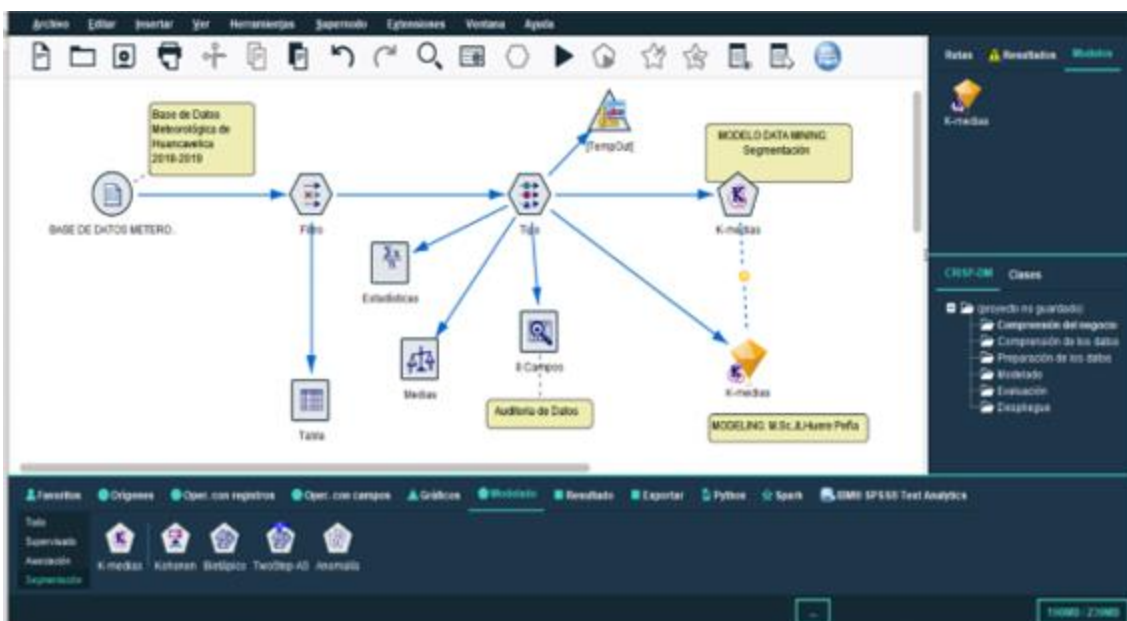


### 3.1.9. Resultado 9

El proceso de Data Mining se inicia con el proceso de modelamiento a partir de los datos planos en formato CSV obtenidos de la consola Vantage Pro, para el cual se utilizó la Metodología CRIP-DM (Cross Industry Standard Process for Data Mining)), que bien incluido en el software IBM – SPSS Modeler que incluye las siguientes etapas: Comprensión del negocio, Compresión de los datos, Preparación de los datos, Modelado, Evaluación y Despliegue, los resultados del modelamiento se muestran en la Figura 22.

**Figura 22**

*Modelo de Data Mining para la determinación de patrones de datos Meteorológicos de la ciudad de Huancavelica 2018-2019.*



A partir del modelo obtenido con la base de datos y siguiendo la metodología CRIP-DM se empleó un algoritmo no supervisado para el cual se usó el modelo de segmentación y utilizando el algoritmo K-means, con el objetivo de obtener cluster a partir de valores meteorológicos diarios en la ciudad de Huancavelica, periodo 2018 – 2019, se obtuvo como resultado la formación de 4 y 10 clústeres, en 10 iteraciones en cada caso, se muestran a continuación los centroides de los clústeres iniciales, las interacciones, los centroides de los clústeres finales y el número de casos en cada clúster.

**Tabla 1**

*Centros de 04 clústeres iniciales de las variables meteorológicas.*

	Clúster			
	1	2	3	4
TempOut	8,65	14,81	8,97	10,75
Rain	,25	,00	,00	1,52
OutHum	63,17	33,94	65,17	62,25
BarmmHg	571,32	550,42	551,40	549,96
SolarRad	2912,00	8716,00	2,00	5806,00
UVIndexmáx	8,70	15,80	,00	14,90
WindSpeedmáx	3,10	4,50	1,30	4,00

**Tabla 2***Historial de iteraciones 04 clúster de las variables meteorológicas*

Iteración	Cambiar en centros de clústeres			
	1	2	3	4
1	444,19	877,09	238,50	358,44
2	46,87	491,98	589,69	133,73
3	16,35	187,03	195,02	101,10
4	2,43	37,15	284,13	30,91
5	3,52	19,52	193,99	23,34
6	8,15	38,73	209,68	23,54
7	27,13	18,82	219,19	11,62
8	58,13	36,01	212,19	6,98
9	68,94	17,33	110,74	21,57
10	85,83	,00	154,27	16,78

a. Se han detenido iteraciones porque se ha realizado el número máximo de iteraciones. Las iteraciones no han podido converger. El cambio de la coordenada máxima absoluta para cualquier centro es 154,273. La iteración actual es 10. La distancia mínima entre los centros iniciales es 2894,087.

**Tabla 3***Centros de 04 clústeres finales de las variables meteorológicas*

	Clúster			
	1	2	3	4
TempOut	9,16	11,08	8,96	9,23
Rain	3,36	1,39	3,00	1,95
OutHum	68,42	50,44	70,25	53,13
BarmmHg	557,32	551,22	557,81	562,34
SolarRad	3685,48	6992,44	2463,40	5176,94
UVIndexmáx	10,66	14,92	7,43	12,44
WindSpeedmáx	3,75	4,44	3,51	4,11

**Tabla 4***Centros de 10 clústeres iniciales de las variables meteorológicas.*

	Clúster									
	1	2	3	4	5	6	7	8	9	10
TempOut	12,49	9,03	8,90	10,46	8,75	12,07	12,65	8,97	9,78	14,81
Rain	,00	5,33	2,03	,51	,00	,00	1,02	,00	,25	,00
OutHum	25,85	76,50	78,00	62,54	49,96	48,54	49,29	65,17	68,46	33,94
BarmmHg	549,05	570,38	552,89	550,73	572,01	548,80	549,79	551,40	571,77	550,42
SolarRad	709,00	3545,00	1639,00	4674,00	5679,00	6662,00	7690,00	2,00	2556,00	8716,00
UVIndexmáx	16,00	10,00	4,30	15,50	10,30	13,00	16,00	,00	7,60	15,80
WindSpeedmáx	5,30	3,60	2,20	4,90	4,50	4,00	5,80	1,30	3,10	4,50

**Tabla 5***Historial de iteraciones a 10 clúster de las variables meteorológicas*

Iteración	Cambiar en centros de clústeres									
	1	2	3	4	5	6	7	8	9	10
1	,00	27,17	170,94	58,60	75,45	5,42	76,01	11,95	137,60	325,76
2	,00	34,47	185,35	61,07	64,71	33,98	65,55	,00	80,38	116,95
3	,00	23,38	49,32	12,17	30,29	63,64	29,68	,00	36,46	,00
4	,00	17,33	110,24	11,79	17,46	14,41	,00	,00	44,94	,00
5	,00	17,32	50,59	6,03	11,51	14,92	,00	,00	35,77	,00
6	,00	23,99	65,22	5,96	5,36	,00	,00	,00	46,48	,00
7	,00	12,03	35,73	,00	,00	,00	,00	,00	26,34	,00
8	,00	,00	47,77	,00	,00	,00	,00	,00	20,34	,00
9	,00	,00	38,70	,00	,00	,00	,00	,00	21,28	,00
10	,00	5,41	12,41	,00	,00	,00	,00	,00	13,53	,00

- Se han detenido iteraciones porque se ha realizado el número máximo de iteraciones. Las iteraciones no han podido converger. El cambio de la coordenada máxima absoluta para cualquier centro es 13,532. La iteración actual es 10. La
- distancia mínima entre los centros iniciales es 708,297.



**Tabla 6***Centros de 10 clústeres finales de las variables meteorológicas*

	Clúster									
	1	2	3	4	5	6	7	8	9	10
TempOut	12,49	9,12	9,04	9,45	9,09	10,51	11,65	6,85	8,92	12,85
Rain	,00	4,03	2,94	3,11	1,12	1,77	1,54	4,19	2,53	,98
OutHum	25,85	69,71	71,53	62,33	48,97	52,59	54,23	75,37	68,24	33,12
BarmmHg	549,05	556,98	558,04	558,24	563,94	552,18	549,76	552,00	559,56	549,41
SolarRad	709,00	3656,02	2404,83	4519,06	5474,77	6534,74	7518,86	6,00	3018,36	8273,33
UVIndexmáx	16,00	10,79	6,94	12,31	12,33	14,43	15,78	,00	8,73	15,97
WindSpeedmáx	5,30	3,71	3,46	3,99	4,12	4,38	4,74	2,00	3,69	4,42

**Tabla 7***Número de casos para 04 y 10 clúster formados con 10 iteraciones en cada caso.**Número de casos en cada clúster*

Clúster	1	114,00
	2	54,00
	3	48,00
	4	142,00
Válidos		358,00
Perdidos		,00

*Número de casos en cada clúster*

Clúster	1	1,00
	2	60,00
	3	24,00
	4	81,00
	5	88,00
	6	35,00
	7	14,00
	8	2,00
	9	47,00
	10	6,00
Válidos		358,00
Perdidos		,00

Posteriormente también a partir del modelo obtenido con la base de datos y siguiendo la metodología CRIP-DM se empleó un algoritmo supervisado en la que se usó como variable dependiente o de supervisión a Rains (que indica un día con lluvia o sin lluvia) para predecir que variable meteorológica presenta mayor influencia en la variable respuesta, para el cual se usó el modelo supervisado de Árbol de decisión, usando en este caso debido a las variables en estudio el Método de crecimiento CRT, con el objetivo de obtener un árbol de decisión y las



influencias de las variables a partir de valores meteorológicos diarios en la ciudad de Huancavelica.

El resultado de este modelo se muestra en la Tabla 12, se obtuvo como resultado la formación de un árbol de decisión con cinco nodos tal como se muestra en la Figura 23 para pronosticar en un día determinado la posibilidad de lluvia en base a la temperatura ambiental y el porcentaje de humedad relativa. En la tabla 13 se muestra el porcentaje de precisión para realizar el pronóstico usando el modelo del árbol de decisión, así como el porcentaje de error del algoritmo.

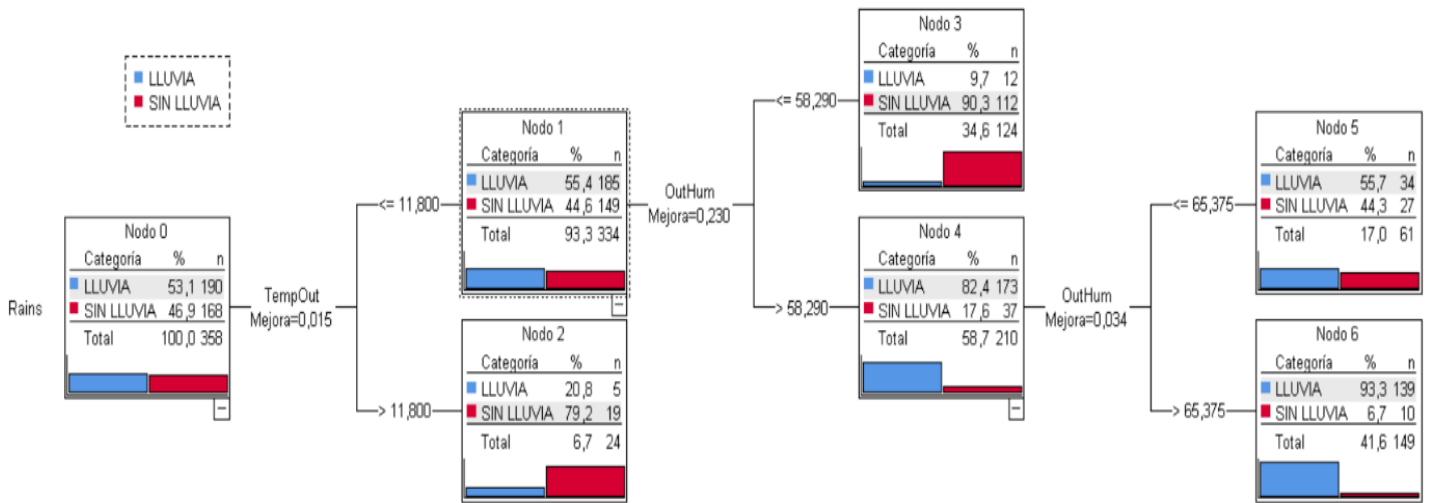
**Tabla 8**

*Resumen del modelo de crecimiento CRT para el árbol de decisiones*

Especificaciones	Método de crecimiento	CRT
	Variable dependiente	Rains
	Variables independientes	OutHum, BarmmHg, SolarRad, UVIndexmáx, WindSpeedmáx, TempOut
	Validación	Ninguna
	Máxima profundidad del árbol	5
	Casos mínimos en nodo padre	100
	Casos mínimos en nodo hijo	50
Resultados	Variables independientes incluidas	TempOut, SolarRad, BarmmHg, UVIndexmáx, OutHum, WindSpeedmáx
	Número de nodos	7
	Número de nodos terminales	4
	Profundidad	3

**Figura 1**

*Árbol de decisión para la variable dependiente precipitación pluviál-lluvias*



**Tabla 9**

*Clasificación del pronóstico y grado de error del árbol de decisión obtenido*

Observado	Pronosticado		
	LLUVIA	SIN LLUVIA	Porcentaje correcto
LLUVIA	173	17	91,1%
SIN LLUVIA	37	131	78,0%
Porcentaje global	58,7%	41,3%	84,9%

Método de crecimiento: CRT  
Variable dependiente: Rains

Error	
Estimación	Desv. Error
,151	,019

Método de crecimiento: CRT  
Variable dependiente: Rains

### 3.2. Discusión

Los datos recogidos mediante una estación meteorológica automática, en principio fueron analizados por una estadística descriptiva, con el fin de cumplir la segunda fase de la metodología CRISP-DM que es la comprensión de los datos (Cruz et al., 2012), usando para ello software de análisis descriptivo como Excel y SPSS statistics en que se observa que los datos muestran comportamientos esperados, pero en algunos casos se presentan anomalías que ocasionan desastres naturales cuando no se toman las medidas de prevención (Acosta & Viale, 2017).

Estas alteraciones o variaciones son evidencias del cambio climático que se presenta (Álvarez Arbulú et al., 2017). Los datos de precipitación y temperatura

son aquellos en que la variación determina otros datos importantes como el caso del cálculo del balance hídrico además de ser datos básicos para la comprensión de la variabilidad meteorológica Primo (Salazar, 2015). El estudio de la variación meteorológica, la determinación de patrones, agrupamiento de características climáticas de acuerdo a las variables y comportamientos temporizados diariamente, mensuales o en intervalos de tiempos predefinidos es importante para dar una interpretación de la realidad que se presenta en una localidad (Norabuena Espinoza, 2014).

La variación meteorológica observada en el periodo de estudio comprendido entre agosto del 2018 a julio del 2019 indica que los datos analizados siguen un patrón regular con algunos datos en algunos meses que salen de los patrones regulares esperados, que es evidencia del cambio climático en la región, el país y el mundo. Este tipo de estudios servirán para recomendar acciones frente a escenarios climáticos adversos los que son conocidos como medidas de adaptación al cambio climático.

A partir de los resultados encontrados, en la presente investigación realizada a partir de datos de los parámetros meteorológicos, muestran comportamientos esperados, pero en otros casos se presentan anomalías que ocasionan desastres naturales cuando no se toman las medidas de prevención (Acosta & Viale, 2017). Esos cambios son evidencias del cambio climático que se presenta hoy en día (Álvarez et al., 2017). Los datos de precipitación y temperatura son aquellos en que la variación determina otros datos importantes como el caso de la determinación de balance hídrico además de ser datos básicos para la comprensión de la variabilidad meteorológica (Primo, 2015).

El estudio de la variación meteorológica y el conocimiento de patrones, es importante para dar una interpretación de la realidad que se presenta en una localidad, así como también poder realizar predicciones a partir de parámetros que puedan influir en otros de manera dependiente (Norabuena, 2014). La variación climática se presenta en cualquier periodo que se analice siendo la precipitación y la temperatura las de mayor irregularidad como en la investigación realizada y están asociadas al crecimiento de las áreas urbanas como indica De Oliveira et al., (2007). Otros estudios que han correlacionado las variables climáticas indican que estos modelos son importantes para predecir el

comportamiento de las variables meteorológicas como también concluye la investigación de Carvajal al. (1999).

El sistema climático está integrado por cinco grandes componentes: la atmósfera, la hidrosfera, la criósfera, la superficie terrestre y la biosfera, y las interacciones entre ellos (Chavez, 2010), es por eso que se hace necesario investigaciones que estudien estas variaciones que se dan en el tiempo para un mejor entendimiento del funcionamiento del clima y descubrimiento de patrones de comportamiento a fin de establecer predicciones. Y las variaciones que muestran mayor irregularidad tiene que ver con el crecimiento de las zonas urbanas respecto a las zonas rurales como también reporta Primo (2015).

Los parámetros como la precipitación y temperatura son irregulares de ellas depende la escasez o abundancia de agua en el planeta por ello también es necesario establecer las relaciones con otras variables meteorológicas tales como la presión atmosférica, radiación solar, humedad relativa, índice de radiación ultravioleta y velocidad del viento tal como como indican Villegas al., (2014) como también indica la investigación de Riquelme (2010).

El estudio detallado del viento también tiene su importancia en la determinación de patrones meteorológicos y la técnica de la rosa de viento permitió determinar comportamientos anuales y mensuales usando software como el WRPLOT a fin de tomar decisiones respecto a movimiento del viento que contiene gases contaminantes y el potencial eólico para una posible implementación de aerogeneradores eólicos (Aragón et al., 2019).

Las técnicas de Data Mining nos ofrecen una variedad de algoritmos algunos de aprendizaje supervisado y otros no supervisados (García, 2020) para la determinación de clúster se usó algoritmos no supervisados obteniéndose grupos significativos de los datos diarios y mensuales tanto en la organización de cuatro y 10 diez grupos con 10 iteraciones en cada caso.

También se usó algoritmos de aprendizaje supervisado denominado árboles de decisión con el objetivo de la predicción de la variable de supervisión de la posibilidad de lluvia obtenida a partir de los días con precipitación pluvial como variable a predecir y las variables temperatura ambiental y humedad relativa

como variables predictoras por ser las de mayor influencia en la variable dependiente u objetivo (Haro, 2020).

### 3.3. Conclusiones

Mediante la aplicación de algoritmos de aprendizaje no supervisado, como la segmentación de datos y los algoritmos k-means, y supervisados como los árboles de decisión, utilizando metodologías computacionales de Data Mining como CRISP-DM, es posible descubrir nuevos conocimientos y comportamientos climáticos con la organización de grupos o clúster a partir de los valores diarios de las variables meteorológicas y realizar predicciones en la ciudad de Huancavelica, estos a su vez que podría usarse también previa comprobación científica en otras ciudades y regiones a fin de tener grupos en las cuales se pueden tomar a partir de estos conocimientos muchas decisiones en campos como la agricultura, la salud, la generación de energías renovables entre otras.

El estudio del comportamiento de las variables meteorológicas de una ciudad y con la cada vez más abundante y creciente cantidad de información obtenida con dispositivos como estaciones meteorológicas automáticas, también es posible realizar predicciones utilizando algoritmos de aprendizaje supervisado da Data Mining como los árboles de decisión en la que se requiere una variable dependiente o a predecir como se realizó en la presente investigación para predecir un día con presencia o ausencia de lluvia a partir de otras variables meteorológicas registradas simultáneamente para su proceso de análisis, obteniendo un alto porcentaje de probabilidad de éxito en la predicción (84.9%).

La técnica de Data Mining para el proceso y análisis de datos tiene una gran cantidad de aplicaciones en las diferentes áreas del conocimiento, en la presente investigación se realizó aplicaciones a los datos meteorológicos, la misma que usando algoritmos de aprendizaje no supervisado como es el k-means se pudo lograr formar clúster utilizando como origen de las mismas las diferentes variables meteorológicas y los 12 meses del año que se recopilaron los datos para su estudio, comprobándose que existen diferencias significativas, mediante pruebas de hipótesis no paramétricas (kolmogorov – Smirnov) debido a que los datos de las variables en estudio no presentan una distribución normal.

## REFERENCIAS

- Alvarez, L., & Borrajero, I. (2016). Caracterización de la marcha anual de fenómenos meteorológicos en Cuba, clasificados según el código de tiempo presente. *Revista Cubana de Meteorología*, XXII(1).
- Aragón, J., Serna, E., & Solano, D. (2019). *Climatological study of the winds for the city of Bogotá in the period 2010-2016*. Cali: Scielo Org.
- Castorena, J. (2018). El uso de herramientas tecnológicas de minería de datos en el análisis de datos climatológicos. *Revista Iberoamericana de la Ciencias Computacionales e Informática*, VII(13). doi:10.23913/reci.v7i13.75
- Cruz, G., Zerquera, R., Morales, A., & Rosete, A. (2012). Estimación de parámetros meteorológicos secundarios en la zona de la Cujae utilizando técnicas de minería de datos. *Revista Cubana de Meteorología*, 18(1).
- Estévez, J. (2008). *Diseño de un Sistema Integrado para el control de calidad de datos de Estaciones Meteorológicas Automáticas*. Universidad de Córdoba. Córdoba: Servicio de Publicaciones de la Universidad de Córdoba. 2010.
- Fundación Española para la Ciencia y la Tecnología. (2004). *Meteorología y Climatología*. Madrid: FECYT.
- García, Z. (2020). *Algoritmos de clasificación supervisados y semi-supervisados: análisis y comparativa*. Madrid: Digital UPM.
- Guerra, I. (2017). *Algoritmos de aprendizaje aplicados a la monitorización meteorológica mediante redes inalámbricas de comunicaciones distribuidas*. Tesis Doctoral, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria.
- Haro, S. (2020). Árbol De Decisión, Aplicación con datos meteorológicos. *VI Congreso Internacional Sectei 2019*. Chimborazo: Knowledge E.
- Hervás, Á. (2020). *Análisis de series temporales no estacionarias en estudio*. Jaén: Universidad de Jaén.
- Khan, S., Muqem, M., & Javed, N. (2016). A Critical Review of Data Mining Techniques in Weather Forecasting. *International Journal of Advanced Research in Computer and Communication Engineering*, V(4). doi:10.17148/IJARCCE.2016.54266
- Mejia, J. (2019). *Herramientas de análisis multivariante predictivo y minería de datos con SPSS Modeler y Statistic*. Guadalajara: Centro Universitario de Ciencias Económico Administrativas de la Universidad de Guadalajara.
- Mendoza, J., Valenzuela, E., Marcos, O., & Gradilla, L. (2020). *Sistema de Información Meteorológica y Climática*. (S. d. Transportes, Ed.) Sanfandila, Mexico.

- Montequín, R. (2002). Técnicas de análisis de datos. (U. d. Oviedo, Ed.)
- Mulatu, W., Bedasa, M., & Terefa, G. (2020). Prediction of Wheat Rust Diseases Using Data Mining Application. *Open Access Library Journal*, VII. doi:10.4236/oalib.1106717
- Oldani, J. (2020). *La Meteorología*. Barcelona: DE VECCHI.
- Organización Meteorológica Mundial. (2010). *Manual del Sistema Mundial de Proceso de Datos y de Predicción*. Ginebra.
- Seda, S., & Todorovic, M. (2020). Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data. *Agricultural Water Management*, CCXXVIII. doi:10.1016/j.agwat.2019.105875
- Serrano, S., Zuleta, D., Moscoso, V., Jácome, P., Palacios, E., & Villacís, M. (2012). Statistical analysis of daily and monthly meteorological data of the metropolitan district of Quito for weather variability and climate change studies. *La Granja*, 16(2), 23-47.
- Soto, C. (2013). *Minería de datos para series temporales en Weka y su aplicación en el pronóstico de precipitaciones*. Universidad Central Marta Abreu.
- Sujay, R., Paresh, C., Mohammad, A., Seyed, M., Nadhir, A., & Zaher, M. (2019). Dew Point Temperature Estimation: Application of Artificial Intelligence Model Integrated with Nature-Inspired Optimization Algorithms. *Water*. doi:10.3390/w11040742
- Tak-chung, F. (2011). A review on time series data mining. (ELSEVIER, Ed.) *Engineering Applications of Artificial Intelligence*, XXIV(1), 164-181. doi:10.1016/j.engappai.2010.09.007
- WMO. (2015). Global aspects: Annex V to the WMO Technical Regulations. *Manual on the Global Observing System, I*.
- World Meteorological Organization. (1993). Guide on the Global Data-Processing System. (305).
- World Meteorological Organization. (2017). Hydrology – From Measurement to Hydrological Information. (WMO, Ed.) *Guide to Hydrological Practices, I*.
- Yusf, N. (2018). Review on weather forecasting using ANN and Data Mining Techniques. *International Journal of Engineering and Techniques*.
- Zamora, T. (2018). *Aplicación de técnicas de minería de datos para pronósticos del sector agrícola*. Facultad de Ingeniería - Escuela de Ingeniería Informática. Valparaiso: Pontificia Universidad Católica de Valparaiso.

## ANEXOS

### Anexo 1. Galería de Fotos

**Foto 01.** Ubicación de la estación meteorológica automatizada. Fuente: Google Maps





**Foto 02.** Estación meteorológica automatizada Vantage Pro CU de Paturpampa Universidad Nacional de Huancavelica con investigador M.Sc. Jorge L. Huere Peña



**Foto 03.** Estación meteorológica automatizada Vantage Pro CU de Paturpampa Universidad Nacional de Huancavelica



**Foto 04.** Consola de la estación meteorológica automatizada Vantage Pro CU de Paturpampa Universidad Nacional de Huancavelica con investigador



Este libro se terminó de publicar en la editorial

**Instituto Universitario  
de Innovación Ciencia y Tecnología Inudi Perú**



ISBN: 978-612-5069-55-9



EDITADA POR  
INSTITUTO  
UNIVERSITARIO  
DE INNOVACIÓN CIENCIA  
Y TECNOLOGÍA INUDI PERÚ