

**UNIVERSIDAD NACIONAL DE HUANCVELICA**  
(Creado por Ley N° 25265)



**ESCUELA DE POSGRADO**

**FACULTAD DE CIENCIAS DE INGENIERIA**  
**UNIDAD DE POSGRADO**

TESIS

**PATRONES DE COMPORTAMIENTO DE DATOS METEOROLÓGICOS  
MEDIANTE TÉCNICAS DE DATA MINING EN  
HUANCVELICA 2018 – 2019**

LÍNEA DE INVESTIGACIÓN: **INVESTIGACIÓN CLIMÁTICA**

PRESENTADO POR:  
**M.Sc. Jorge Luis, HUERE PEÑA**

PARA OPTAR EL GRADO ACADÉMICO DE DOCTOR EN  
**CIENCIAS AMBIENTALES**

HUANCVELICA – PERÚ  
**2021**



UNIVERSIDAD NACIONAL DE HUANCVELICA  
(Creado por Ley N° 25265)



**ESCUELA DE POSGRADO**

**FACULTAD DE CIENCIAS DE INGENIERIA**  
**UNIDAD DE POSGRADO**

(APROBADO CON RESOLUCIÓN N° 736-2005-ANR)

"Año del Bicentenario del Perú: 200 años de Independencia"

### ACTA DE SUSTENTACIÓN DE TESIS

Ante el Jurado conformado por los docentes: **Dr. Manuel Emiliano ESTEVES PAIRAZAMAN**, **Dr. Victor Guillermo SANCHEZ ARAUJO**, **Dr. William Herminio SALAS CONTRERAS**

**Asesor: Dr. Jose Luis GAVE CHAGUA**

De conformidad al Reglamento Único de Grados y Títulos de la Universidad Nacional de Huancavelica, aprobado mediante Resolución N° 330-2019-CU-UNH, y modificado con Resolución N° 552-2021-CU-UNH, y la Directiva de la Sustentación Síncrona de Tesis de los Estudiantes de Maestría y Doctorado de las Unidades de Posgrado de las Facultades Integrantes de la Universidad Nacional de Huancavelica en el Marco al estado de emergencia covid 19, aprobado mediante Resolución Directoral N° 340-2020-EPG-RUNH.

El candidato al **GRADO DE DOCTOR EN CIENCIAS AMBIENTALES**

Don, **Jorge Luis HUERE PEÑA**, procedió a sustentar su trabajo de Investigación titulado "**PATRONES DE COMPORTAMIENTO DE DATOS METEOROLÓGICOS MEDIANTE TÉCNICAS DE DATA MINING EN HUANCVELICA 2018 - 2019**".

Luego de haber absuelto las preguntas que le fueron formulados por los Miembros del Jurado, se dio por concluido al ACTO de sustentación, realizándose la deliberación y calificación, resultando:

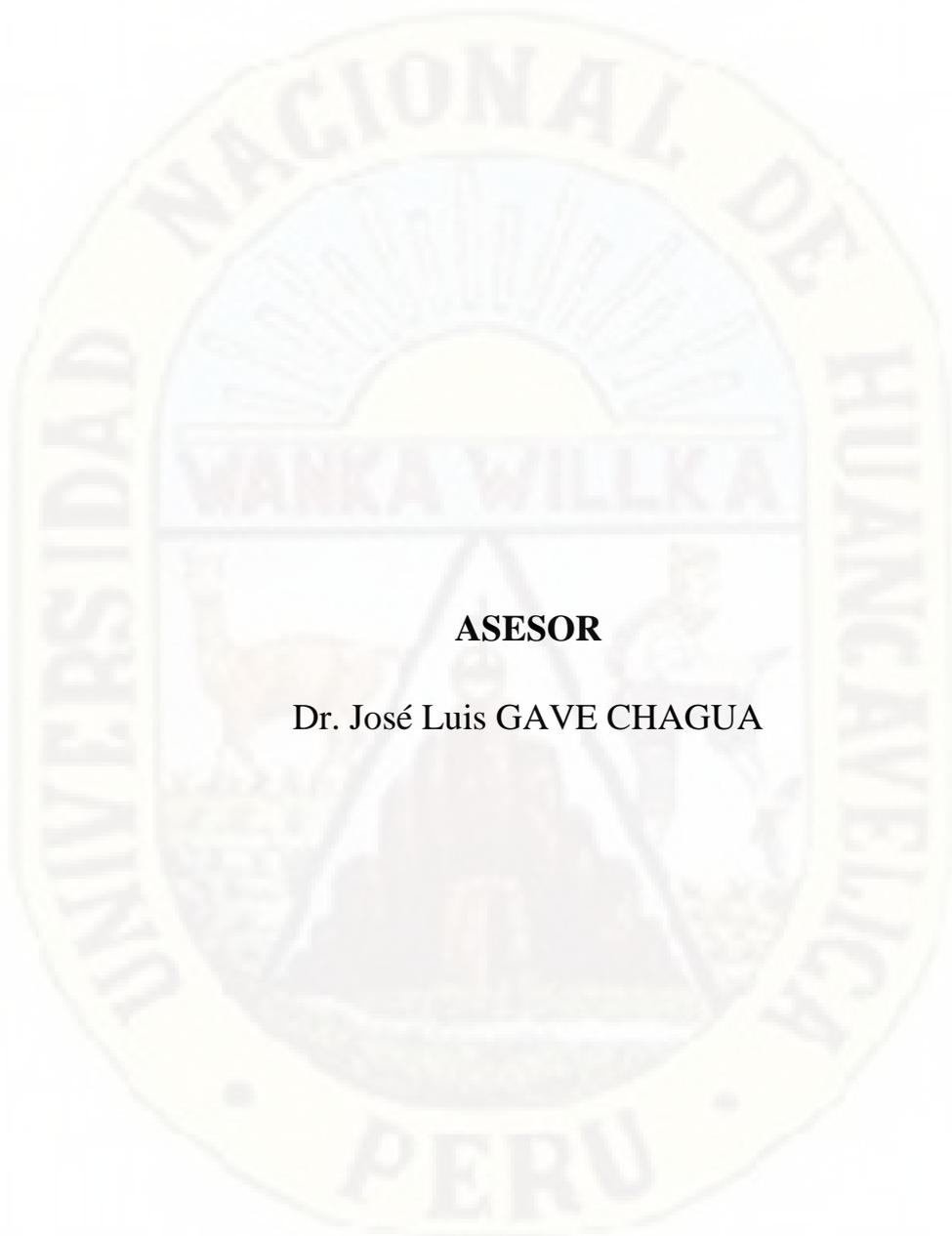
Con el calificativo de:

Y para constancia se extiende la presente ACTA, en la ciudad de Huancavelica, a los veintiséis días del mes de octubre del año 2021.

  
.....  
**Dr. Manuel Emiliano ESTEVES PAIRAZAMAN**  
Presidente del Jurado.

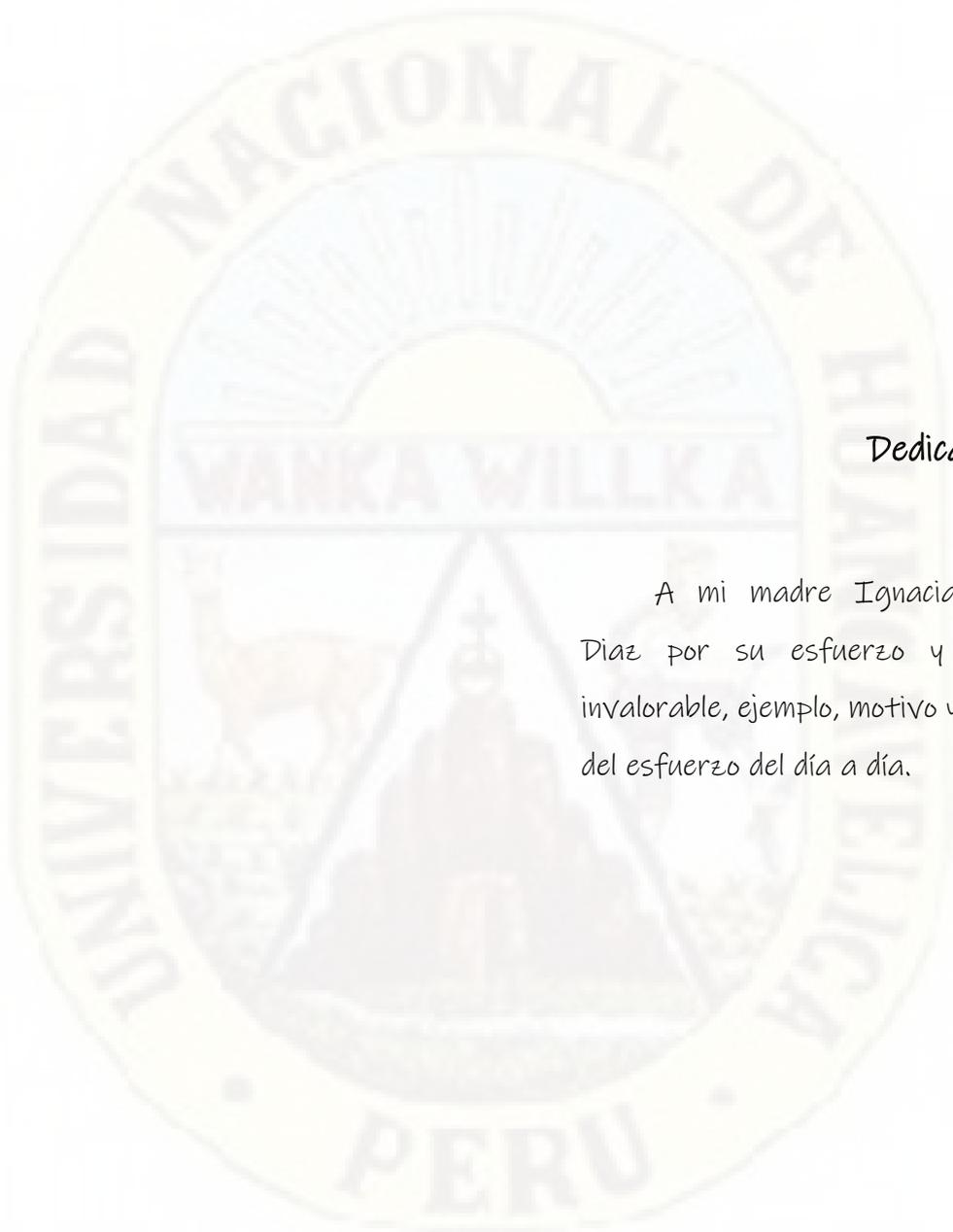
  
.....  
**Dr. Victor Guillermo SANCHEZ ARAUJO**  
Secretario del Jurado

  
.....  
**Dr. William Herminio SALAS CONTRERAS**  
Vocal del Jurado



**ASESOR**

**Dr. José Luis GAVE CHAGUA**



## Dedicatoria

A mi madre Ignacia Peña Diaz por su esfuerzo y apoyo invaluable, ejemplo, motivo y razón del esfuerzo del día a día.

## Resumen

Una de las fases del descubrimiento de conocimiento de bases de datos (KDD), comprende las técnicas de Data Mining o minería de datos, en este marco la presente investigación tuvo como objetivo determinar patrones de comportamiento de datos obtenidos mediante éstas técnicas, de las variables meteorológicas en la ciudad de Huancavelica como son: la temperatura ambiental, presión atmosférica, humedad atmosférica, velocidad del viento, radiación solar, radiación ultra violeta y precipitación pluvial durante el periodo 2018 y 2019, 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 mediante el Software SPSS Modeler que en la ruta de sus procedimientos, incluye las etapas de esta metodología, 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 claves:** Variables meteorológicas, patrones meteorológicos, comportamiento meteorológico, Data Mining.

## Abstract

One of the phases of the discovery of knowledge of databases (KDD), includes the techniques of Data Mining or *Minería de datos*, in this framework the present research aimed to determine patterns of behavior of data obtained through these techniques, of the variables weather conditions in the city of Huancavelica such as: environmental temperature, atmospheric pressure, atmospheric humidity, wind speed, solar radiation, ultraviolet radiation and rainfall during the period 2018 and 2019, using an automated weather station from the Weather company. Link, DAVIS Brand, Vantage Pro Model and a console for the storage of Vantage Pro data in plain text and that were later processed, described and analyzed using SPSS Statistical software and WRPLOT in the particular case of the variable wind direction and for the determination of behaviors and patterns, the CRISP-DM methodology was used through the Software SPSS Modeler that in the route of its procedures, includes the stages of this methodology, the results obtained were clustering of meteorological variables with unsupervised learning algorithms and predictions of the rainfall variable with supervised learning algorithms obtaining 84.9% of probabilities 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.

# Índice General

RESUMEN .....	V
ABSTRACT .....	VI
ÍNDICE GENERAL.....	VII
ÍNDICE DE TABLAS .....	X
ÍNDICE DE FIGURAS .....	XI
INTRODUCCIÓN .....	XIII
<b>CAPÍTULO I EL PROBLEMA.....</b>	<b>1</b>
<b>1.1.PLANTEAMIENTO DEL PROBLEMA.....</b>	<b>1</b>
<b>1.2.FORMULACIÓN DEL PROBLEMA .....</b>	<b>2</b>
<i>1.2.1. Problema General.....</i>	<i>2</i>
<i>1.2.2. Problemas Específicos .....</i>	<i>2</i>
<b>1.3.OBJETIVOS DE LA INVESTIGACIÓN.....</b>	<b>2</b>
<i>1.3.1. Objetivo general.....</i>	<i>2</i>
<i>1.3.2. Objetivos Específicos .....</i>	<i>2</i>
<b>1.4.JUSTIFICACIÓN .....</b>	<b>3</b>
<b>CAPÍTULO II MARCO TEÓRICO .....</b>	<b>4</b>
<b>2.1.ANTECEDENTES DE LA INVESTIGACIÓN .....</b>	<b>4</b>
<i>2.1.1. Antecedentes Internacionales.....</i>	<i>4</i>
<i>2.1.2. Antecedentes Nacionales .....</i>	<i>5</i>
<i>2.1.3. A nivel local.....</i>	<i>6</i>
<b>2.2.BASES TEÓRICAS .....</b>	<b>6</b>
<i>2.2.1. Cambio climático y variabilidad climática .....</i>	<i>6</i>
<i>2.2.2. Meteorología, tiempo meteorológico y clima .....</i>	<i>7</i>
<i>2.2.3. Patrones de comportamiento de datos meteorológicos .....</i>	<i>7</i>
<i>2.2.4. Algoritmos computacionales y procesamiento de datos meteorológicas .....</i>	<i>8</i>
<b>2.3.MARCO CONCEPTUAL.....</b>	<b>9</b>
<i>2.3.1. Aspectos generales .....</i>	<i>9</i>
<i>a) Meteorología.....</i>	<i>9</i>

b) Descripción de la atmósfera .....	9
c) Organización Meteorológica Mundial (OMM) .....	10
d) Fuentes de información meteorológica .....	10
e) Calidad y validación de datos meteorológicos .....	11
2.3.2. Variables meteorológicas y patrones de comportamiento .....	13
a) Variables meteorológicas, concepto, instrumento y unidades de medida. ....	13
b) Sistemas de observación de variables meteorológicas .....	16
c) Patrones de comportamiento meteorológico .....	18
d) Técnicas para la determinación de patrones de comportamiento. ....	19
e) Metodologías, Técnicas y aplicaciones de Data Mining a datos meteorológicos. ....	20
f) Data Mining. Algoritmos de aprendizaje supervisado y no supervisado .....	22
g) Análisis de series temporales .....	22
h) Clasificación de series temporales en WEKA .....	23
i) Análisis de series temporales con RStudio software .....	24
j) Data Mining para Series Temporales .....	24
k) Análisis de datos y modelos de predicción mediante IBM SPSS Modeler software .....	25
l) Técnica del Árbol de decisión y su aplicación a datos meteorológicos. ....	26
2.4.MARCO FILOSÓFICO .....	26
2.4.1. Encuadramiento Epistemológico del Problema .....	26
2.4.2. El Pragmatismo .....	26
2.4.3. Racionalismo .....	27
2.4.4. El Empirismo .....	27
2.4.5. Intelectualismo: .....	28
2.4.6. El Apriorismo .....	28
2.4.7. El Objetivismo .....	29
2.4.8. El Realismo .....	29
2.4.9. El Fenomenalismo .....	30
2.4.10. ¿Qué es un Metalenguaje? .....	30
2.4.11. Sistema Procesador de Operaciones Metalingüísticas .....	30
2.5.FORMULACIÓN DE HIPÓTESIS .....	31
2.5.1. Hipótesis General .....	31
2.5.2. Hipótesis Específicas .....	31
2.6.IDENTIFICACIÓN DE VARIABLES .....	31
2.6.1. Variable I .....	31
2.7.DEFINICIÓN OPERATIVA DE VARIABLES E INDICADORES .....	32
2.7.1. Variable I: .....	32

<b>CAPÍTULO III METODOLOGÍA DE LA INVESTIGACIÓN .....</b>	<b>33</b>
3.1.TIPO DE INVESTIGACIÓN .....	33
3.2.NIVEL DE INVESTIGACIÓN .....	33
3.3.MÉTODOS DE INVESTIGACIÓN.....	34
3.4.DISEÑO DE INVESTIGACIÓN .....	34
3.5.POBLACIÓN, MUESTRA Y MUESTREO .....	35
3.5.1. <i>Población</i> .....	35
3.5.2. <i>Muestra</i> .....	35
3.5.3. <i>Muestreo</i> .....	35
3.6.TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS.....	36
3.6.1. <i>Técnicas</i> .....	36
3.6.2. <i>Los instrumentos</i> .....	36
3.7.TÉCNICAS DE PROCESAMIENTO Y ANÁLISIS DE DATOS.....	37
3.8.DESCRIPCIÓN DE LA PRUEBA DE HIPÓTESIS .....	38
<b>CAPÍTULO IV PRESENTACIÓN DE RESULTADOS.....</b>	<b>40</b>
4.1.PRESENTACIÓN E INTERPRETACIÓN DE DATOS .....	40
4.2.DISCUSIÓN DE RESULTADOS .....	59
4.3.PROCESO DE PRUEBA DE HIPÓTESIS.....	62
4.3.1. <i>Pruebas de Normalidad</i> .....	62
4.3.2. <i>Pruebas de Hipótesis</i> .....	67
CONCLUSIONES.....	72
RECOMENDACIONES.....	73
REFERENCIAS BIBLIOGRÁFICAS.....	74
ANEXOS .....	77
MATRIZ DE CONSISTENCIA .....	78
INSTRUMENTO DE RECOLECCIÓN DE DATOS .....	79
BASE DE DATOS.....	80
ARTÍCULO CIENTÍFICO.....	90

## Índice de Tablas

<b>Tabla 1</b> Límites físicos de diferentes variables meteorológicas.....	12
<b>Tabla 2</b> Tolerancias propuestas en función del tiempo transcurrido entre medidas consecutivas.....	12
<b>Tabla 3</b> Instrumentos de variables meteorológicas .....	16
<b>Tabla 4</b> Operacionalización de las variables meteorológicas de la ciudad de Hvca. ....	32
<b>Tabla 5</b> Centros de 04 clústeres iniciales de las variables meteorológicas.....	54
<b>Tabla 6</b> Historial de iteraciones 04 clúster de las variables meteorológicas .....	54
<b>Tabla 7</b> Centros de 04 clústeres finales de las variables meteorológicas.....	55
<b>Tabla 8</b> Centros de 10 clústeres iniciales de las variables meteorológicas.....	55
<b>Tabla 9</b> Historial de iteraciones 10 clúster de las variables meteorológicas .....	56
<b>Tabla 10</b> Centros de 10 clústeres finales de las variables meteorológicas.....	56
<b>Tabla 11</b> Número de casos para 04 y 10 clúster formados con 10 iteraciones en cada caso.....	57
<b>Tabla 12</b> Resumen del modelo de crecimiento CRT para el árbol de decisiones .....	58
<b>Tabla 13</b> Clasificación del pronóstico y grado de error del árbol de decisión obtenido	59
<b>Tabla 14</b> Prueba de Normalidad de las variables meteorológicas en estudio.....	63
<b>Tabla 15</b> Tabla de ANOVA respecto a cuatro clústeres formados mediante la técnica de Data Mining denominada segmentación con el algoritmo no supervisado k-means.....	68
<b>Tabla 16</b> Tabla de ANOVA respecto a diez clústeres formados mediante la técnica de Data Mining denominada segmentación con el algoritmo no supervisado k-means.....	69
<b>Tabla 17</b> Tabla de la prueba de hipótesis respecto las variables meteorológicas en meses, ciudad de Hvca., 2018 – 2019.....	70

## Índice de Figuras

<b>Figura 1</b> Sistema de observación meteorológica .....	17
<b>Figura 2</b> Estación Sinóptica Meteorológica (ESIME) .....	18
<b>Figura 3</b> Etapas KDD. Data Mining es el núcleo de un proceso de descubrimiento de conocimiento.....	19
<b>Figura 4</b> Esquema comparativo de las tareas generales y fases de las Metodologías de Data Mining SEMMA y CRIP-DM .....	21
<b>Figura 5</b> Diseño de la investigación mediante SPSS Modeler para determinar Patrones Meteorológicos en la ciudad de Huancavelica 2018-2019.....	34
<b>Figura 6</b> Estación Meteorológica Automatizada Weather Link, Marca DAVIS, Modelo Vantage Pro .....	37
<b>Figura 7</b> Software usado para la descripción, análisis y Data Mining de los datos meteorológicos .....	38
<b>Figura 8</b> Temperatura Ambiental de la ciudad de Huancavelica, periodo 2018-2019. .	41
<b>Figura 9</b> Precipitación Pluvial de la ciudad de Huancavelica, periodo 2018-2019. ....	42
<b>Figura 10</b> Humedad relativa de la ciudad de Huancavelica, periodo 2018-2019.....	43
<b>Figura 11</b> Presión atmosférica mensual de la ciudad de Huancavelica, periodo 2018-2019. ....	44
<b>Figura 12</b> Variación horaria de la Presión atmosférica en la ciudad de Huancavelica, periodo 2018-2019.....	44
<b>Figura 13</b> Radiación Solar de la ciudad de Huancavelica por meses, periodo 2018-2019. ....	45
<b>Figura 14</b> Radiación Solar de la ciudad de Huancavelica por horas, periodo 2018-2019. ....	46
<b>Figura 15</b> Radiación Ultravioleta de la ciudad de Huancavelica por meses, periodo 2018-2019. ....	47
<b>Figura 16</b> Radiación Ultravioleta de la ciudad de Huancavelica por horas, periodo 2018-2019. ....	47
<b>Figura 17</b> Velocidad promedio del viento mensual a las 7, 13 y 19 horas en la ciudad de Huancavelica, periodo 2018-2019.....	48
<b>Figura 18</b> Rosa de viento: “blowing from” en la ciudad de Huancavelica- periodo 2018-2019. ....	49

<b>Figura 19</b> Rosa de viento: “blowing to” en la ciudad de Huancavelica- periodo 2018-2019. ....	50
<b>Figura 20</b> Rosa de viento: “blowing from” en la ciudad de Huancavelica por meses 2018-2019 .....	51
<b>Figura 21</b> Rosa de viento: “blowing to” en la ciudad de Huancavelica por meses 2018-2019 .....	52
<b>Figura 22</b> Modelo de Data Mining para la determinación de patrones de datos Meteorológicos de la ciudad de Huancavelica 2018-2019.....	53
<b>Figura 23</b> Árbol de decisión para la variable dependiente precipitación pluvial-lluvias	58
<b>Figura 24</b> Curva de normalidad de la variable temperatura ambiental promedio diario .....	64
<b>Figura 25</b> Curva de normalidad de la variable humedad relativa promedio diario .....	64
<b>Figura 26</b> Curva de normalidad de la variable presión atmosférica promedio diario... 65	
<b>Figura 27</b> Curva de normalidad de la variable Radiación solar acumulada diario .....	65
<b>Figura 28</b> Curva de normalidad de la variable índice de radiación ultravioleta máxima diaria.....	66
<b>Figura 29</b> Curva de normalidad de la variable velocidad del viento máxima diaria .....	66
<b>Figura 30</b> Curva de normalidad de la variable precipitación pluvial acumulada diaria	67
<b>Figura 31</b> Instrumento de medición de variables meteorológicas (EMA) Estación Meteorológica Automatizada .....	79
<b>Figura 32</b> Ubicación de la estación meteorológica automatizada. Fuente: Google Maps .....	92
<b>Figura 33</b> Estación meteorológica automatizada Vantage Pro CU de Paturpampa Universidad Nacional de Huancavelica con investigador M.Sc. Jorge L. Huere Peña .....	92
<b>Figura 34</b> Estación meteorológica automatizada Vantage Pro CU de Paturpampa Universidad Nacional de Huancavelica .....	93
<b>Figura 35</b> Consola de la estación meteorológica automatizada Vantage Pro CU de Paturpampa Universidad Nacional de Huancavelica con investigador.....	93

## **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).

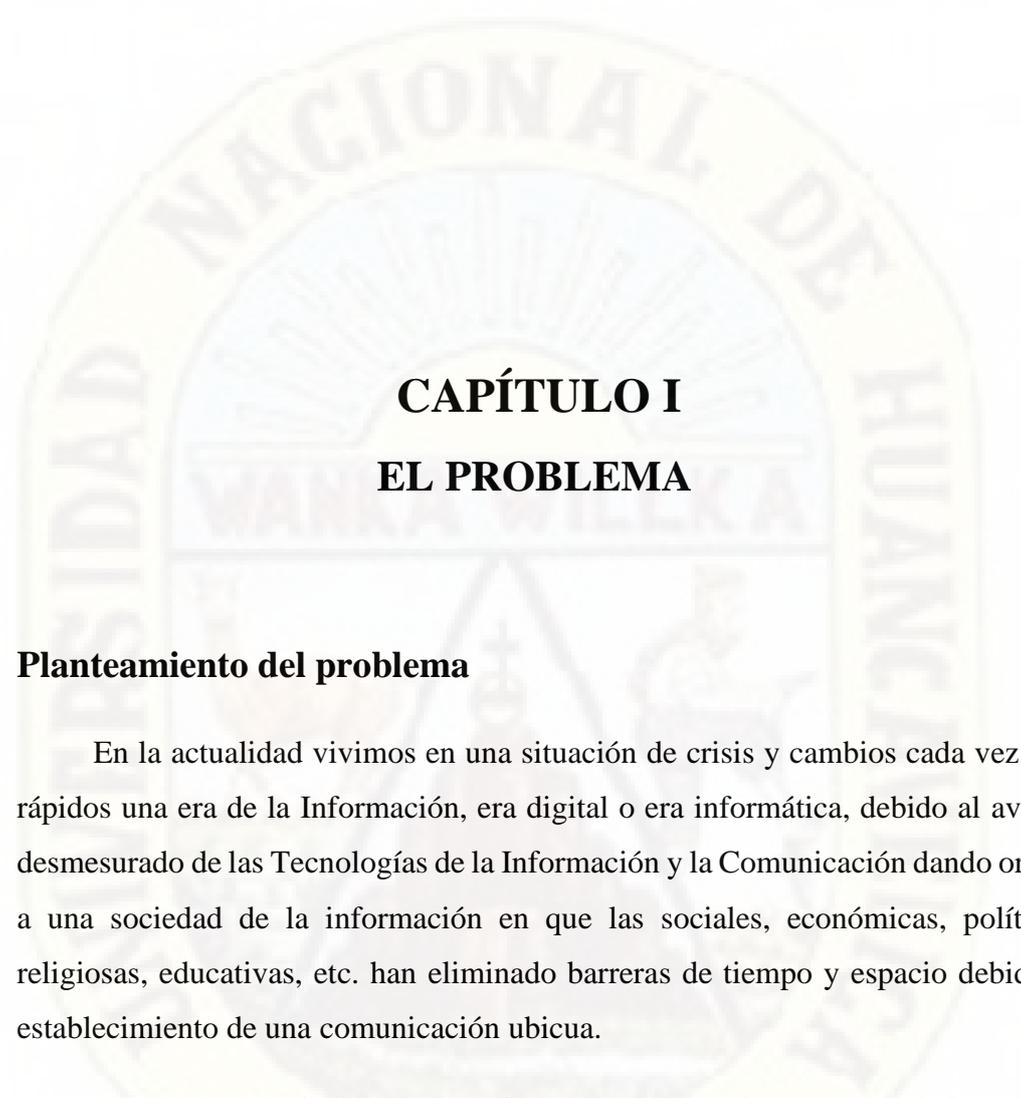
Es así que el estudio de la variación de los meteoros que se presentan en la atmósfera terrestre en periodos determinados es de utilidad para los estudios del clima a nivel regional y local (Páez Martínez, Espitia Barrera, & Cárdenas Castro, 2013), como lo han sido desde tiempos pasados en los que su 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, & Katzschner, 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 centran 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

## EL PROBLEMA

### 1.1. Planteamiento del problema

En la actualidad vivimos en una situación de crisis y cambios cada vez más rápidos una era de la Información, era digital o era informática, debido al avance desmesurado de las Tecnologías de la Información y la Comunicación dando origen a una sociedad de la información en que las sociales, económicas, políticas, religiosas, educativas, etc. han eliminado barreras de tiempo y espacio debido al establecimiento de una comunicación ubicua.

Esta nueva sociedad de la información y con los recursos informáticos emergentes y existentes disponen hoy en día de un creciente conjunto de datos (*Big Data*) de diversos tipos: datos de tipo geoespacial, movimientos bancarios, educativos, empresariales, meteorológicos, etc. cuya complejidad debido a su tamaño y velocidad de crecimiento dificultan su recolección, procesamiento y administración mediante herramientas convencionales denominados *Big Data*.

## **1.2. Formulación del problema**

### **1.2.1. Problema General**

¿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. Problemas Específicos**

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

### **1.3.1. 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.

### **1.3.2. Objetivos Específicos**

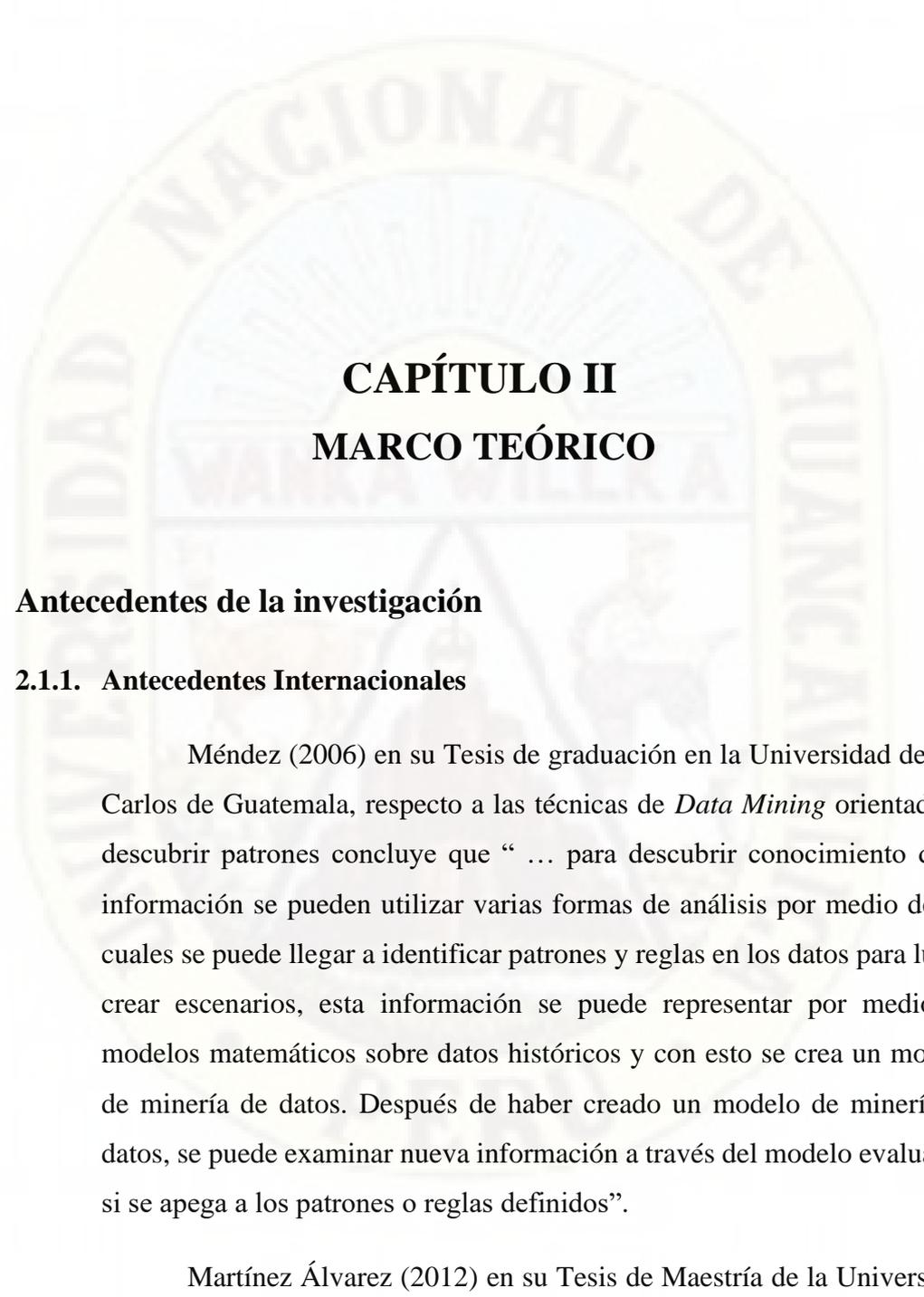
- a) Determinar si existen diferencias significativas entre clúster formados mediante técnicas de Data Mining, a partir de valores meteorológicos diarios en la ciudad de Huancavelica, 2018 – 2019.
- b) Determinar si existe diferencia significativa entre las variables meteorológicas procesados mediante técnicas de Data Mining y los diferentes meses del año en la ciudad de Huancavelica, 2018 – 2019.

#### **1.4. Justificación**

El presente trabajo de investigación nos permitirá descubrir las relaciones inherentes entre las nuevas técnicas de procesamiento de datos aplicados a datos meteorológicos, como es el caso de los *Data Mining* orientados a descubrir nuevos patrones de comportamiento a partir de estas técnicas y algoritmos computacionales implícitos en ellas y que posteriormente puedan contribuir a la toma de decisiones más acertadas que apoyan a la gestión ambiental y con conocimiento de sus patrones y tendencias.

Las técnicas de procesamiento de datos convencionales no son las únicas fuentes para la 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 de tipo meteorológicos existentes y crecientes en el tiempo, hace necesario el desarrollo de métodos y técnicas modernas que permitan su procesamiento, análisis y el descubrimiento del comportamiento de datos meteorológicos a fin de tomar las decisiones más acertadas y prevenir riesgos de tipo ambiental basadas en la incertidumbre y las probabilidades de ocurrencia cuando se realizan predicciones basado en estas técnicas computacionales en el monitoreo ambiental.



## **CAPÍTULO II**

### **MARCO TEÓRICO**

#### **2.1. Antecedentes de la investigación**

##### **2.1.1. Antecedentes Internacionales**

Méndez (2006) en su Tesis de graduación en la Universidad de San Carlos de Guatemala, respecto a las técnicas de *Data Mining* orientados a descubrir patrones concluye que “ ... para descubrir conocimiento de la información se pueden utilizar varias formas de análisis por medio de las cuales se puede llegar a identificar patrones y reglas en los datos para luego crear escenarios, esta información se puede representar por medio de modelos matemáticos sobre datos históricos y con esto se crea un modelo de minería de datos. Después de haber creado un modelo de minería de datos, se puede examinar nueva información a través del modelo evaluando si se apega a los patrones o reglas definidos”.

Martínez Álvarez (2012) en su Tesis de Maestría de la Universidad de Chile respecto a las técnicas de *Data Mining* concluye que “ ... con ayuda de minería de datos se logró diseñar nuevas métricas para el proceso e identificar a los clientes y servicios más críticos, lo que permite llegar a valores más exactos de los ingresos perdidos en cada segmento, y aplicar

estrategias diferenciadas para hacer el cobro a sus clientes, además el trabajo realizado permitió una reducción del tiempo de obtención de los indicadores en un 78%, logrando además estimar los ingresos perdidos mensualmente por servicios no facturados en un monto de MM \$ 210, con un error de la estimación menor al 5%. Se espera que, con ayuda de este estudio, la empresa pueda tomar decisiones informadas y mejorar su capacidad de control del proceso de provisión de servicios privados, con el fin de regularizar su flujo de ingreso mensual”.

Cedano (2015) en su tesis para optar el grado de magister en la Universidad Tecnológica Nacional de México, respecto a la obtención de patrones mediante técnicas de *Data Mining* concluye que “... la toma de decisiones implementada con inteligencia de negocios, a través de herramientas de minería de datos, contribuirá de gran manera a una mejor planeación en el área administrativa, docente y psicopedagógica, para evitar el rezago estudiantil y apoyar en todo momento al alumnado ...”.

Ortiz y Olivares (2015) en la Tesis para optar el grado de magister en la Pontificia Universidad Javeriana, respecto a la variable Gestión Ambiental concluye que “...El sector hotelero bogotano, certificado en las normas NTS-TS-002:2006 e ISO 14001:2004 y representado por 32 hoteles evaluados y adscritos al Ministerio de Comercio, Industria y Turismo, mostró conocimiento en los principales impactos ambientales y el modelo de gestión ambiental nacional. La norma NTS-TS-002:2006 mostró ser una herramienta eficiente para el mejoramiento de ciertos parámetros de gestión ambiental; sin embargo, las mejoras son dependientes de las necesidades y el énfasis que cada establecimiento le quiera dar a la norma”.

### **2.1.2. Antecedentes Nacionales**

Villafuerte (2015) en su tesis realizado en la Universidad Nacional José Faustino Sánchez Carrión, concluye que “... durante la investigación se encontró falta de interés y resistencia en implantar la gestión ambiental para preservar el medio ambiente de parte de los gerentes en su mayoría, no

capacitan a sus colaboradores en temas ambientales, así mismo los huéspedes no son concientizados, por otro lado los productos químicos de limpieza, son utilizados de manera inadecuada, se encontró además, en casi todos los hoteles, el uso de los recursos naturales de manera indiscriminada y el reciclaje no se realiza en ninguna de sus áreas. Por ende, las políticas, las buenas prácticas ambientales, la cultura ambiental, ecoeficiencia y preservación no son de interés de los gerentes .... “.

Gutarra (2006) en la tesis realizada en la Universidad Agraria La Molina, busca “...presentar el paso a paso de la implementación del sistema de gestión ambiental bajo la norma ISO 14001 en esta empresa dedicada a dar servicios de perforación a minas. La metodología empleada para esta implementación fue la propuesta por la Casa Matriz, basada en la misma norma, comenzando con la obtención del compromiso de la empresa hasta la misma certificación, sin dejar de lado el mejoramiento continuo del sistema”.

### **2.1.3. A nivel local**

A nivel local se realizaron una búsqueda de información en las universidades públicas y privadas, instituciones dedicadas al área del medio ambiente y Meteorología y no se encontraron referencias parecidas al presente estudio.

## **2.2. Bases teóricas**

### **2.2.1. Cambio climático y variabilidad climática**

Cambio climático se define como una modificación del clima duradera y significativa con respecto a un historial climático a escala regional y global y sobre variados parámetros meteorológicos. Sus causas pueden ser la variabilidad natural del clima y antropogénicas a través de su alteración sobre la composición química atmosférica, debido a la mayor presencia de Gases de Efecto Invernadero en todo el mundo (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, y otros, 2012).

La variabilidad climática consiste en aquellos cambios sobre el clima que dependen de condiciones atmosféricas extremas que exceden a los promedios estándares. Los fenómenos que producen esos contrastes pueden ser frentes fríos muy organizados, células estacionarias secas, huracanes, perturbaciones tropicales y células con una humedad desproporcionada (Serrano, y otros, 2012).

### **2.2.2. Meteorología, tiempo meteorológico y clima**

Meteorología, tiempo o tiempo meteorológico y clima, aparentemente se refieren al mismo evento, pero no lo es. **Meteorología** es el estudio de la atmosfera y de los mecanismos que producen cambios en la misma. El concepto de **tiempo meteorológico** o simplemente tiempo, hace referencia a los mecanismos de cambio estudiados en periodos cortos de tiempo. A escala de una vida humana hablamos de un momento específico, tiempo, es decir estado de la atmosfera en un momento específico y sus implicaciones, por ejemplo, en este contexto se suele referir al tiempo en términos de luminosidad, lluvioso, factores aéreos y factores eléctricos, fundamentalmente, términos que engloban entre otros a los hidrometeoros, lluvia, nieve, granizo, entre otros. Por otro lado, el concepto de **clima** está asociado cuando se trata de estudiar variaciones de tiempo en un largo recorrido (años) de tiempo meteorológico (Guerra, 2017).

### **2.2.3. Patrones de comportamiento de datos meteorológicos**

En los años recientes, se han estado desarrollando modelos climáticos y patrones de comportamiento a partir de datos meteorológicos a alta resolución, usados para estudiar los cambios a largo plazo, sobre todo el actual cambio climático. Sin embargo, hay que ser cuidadosos en este sentido: el clima es el promedio estadístico a largo plazo de los datos meteorológicos obtenidos en estaciones meteorológicas ubicadas en una

zona determinada que presentan características similares y que definen un clima determinado. Esto se hace en todos los tipos climáticos de todo el mundo. Pero estos tipos climáticos no pueden condensarse en determinados modelos o patrones de comportamiento porque las variaciones a largo plazo de los mismos deben ser obtenidas a posteriori de dichas variaciones producidas a largo plazo.

Dicho en otros términos: la información meteorológica obtenida en multitud de estaciones meteorológicas de todo el mundo sirve, de manera inductiva, para establecer las características climáticas con sus variantes en toda la superficie terrestre y una vez que las obtenemos 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 escalas distintas, solo se usará como probabilidades o patrones de comportamiento posibles.

#### **2.2.4. Algoritmos computacionales y procesamiento de datos meteorológicas**

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, y otros, 2019) (Aragón, Serna, & Solano, 2019).

## **2.3. Marco conceptual**

### **2.3.1. Aspectos generales**

#### **a) Meteorología**

La Meteorología se define como la ciencia encargada del estudio de la atmósfera, de sus propiedades y de los fenómenos que en ella tienen lugar, conocidos como meteoros. El estudio de la atmósfera se basa en el conocimiento de una serie de magnitudes, o variables meteorológicas, como son la temperatura, la presión atmosférica, radiación solar, velocidad del viento y otras variables, las cuales varían tanto en espacio como en tiempo (Fundación Española para la Ciencia y la Tecnología, 2004).

#### **b) Descripción de la atmósfera**

La atmósfera terrestre es la capa gaseosa que envuelve la Tierra, se adhiere a ella debido a la acción de la gravedad y está compuesta de unas masas de Aire que son una mezcla de gases. Los principales gases que componen el aire terrestre son: el nitrógeno ( $N_2$ ) al 78.084%, oxígeno ( $O_2$ ) al 20.95%, argón (Ar) al 0.93%, anhídrido carbónico ( $CO_2$ ) al 0.03%, Neón (Ne) al 0.02%, Helio (He) al 0.0005%, Metano ( $CH_4$ ) al 0.001%, Criptón (Kr) al 0.001% y vapor de agua, este último de gran importancia para la meteorología (Oldani, 2020). Es importante precisar que la concentración de estos gases varía con la altura. Cuando se produce una transición entre dos masas de aire de distintas densidades y temperaturas aparece lo que se denomina un Frente. Estos frentes además de traer la precipitación producen cambios en la presión de

la atmosfera, la temperatura, vientos y otros fenómenos (Guerra, 2017) (Fundación Española para la Ciencia y la Tecnología, 2004).

**c) Organización Meteorológica Mundial (OMM)**

El Sistema Mundial de Proceso de Datos y de Predicción (SMPDP) tiene como principal objetivo de preparar y poner a disposición de los Miembros, de la manera más económica y efectiva, análisis meteorológicos y productos de predicción y su estructura organizativa está conformado por los Centros Meteorológicos Mundiales (CMM), los Centros Meteorológicos Regionales Especializados (CMRE) y los Centros Meteorológicos Nacionales (CMN) en cada país, en algunos casos estos centros se constituyen como uno solo y se encarga de llevar a cabo simultáneamente las funciones de todos ellos (Organización Meteorológica Mundial, 2010).

**d) Fuentes de información meteorológica**

Las fuentes de información meteorológica permiten la observación y medida de las variables y fenómenos meteorológicos, estas fuentes de información se encuentran en miles de estaciones meteorológicas ubicadas sobre la superficie de la tierra en posiciones fijas, pero también sobre el mar (barcos, buques y otros) y a distintas alturas de la atmósfera (satélites y aviones) aprovechando que todos ellos disponen de instrumentos meteorológicos a bordo. El uso que se puede dar a la información que proporcionan todas estas fuentes de observación es muy amplio y variado: desde el mero registro temporal en estaciones concretas, a la elaboración de predicciones meteorológicas. En cualquier caso, los centros meteorológicos centralizan la información por áreas, la procesan, controlan su calidad, y la distribuyen a los usuarios que puedan necesitarla para investigaciones y estudio de la atmósfera.

e) **Calidad y validación de datos meteorológicos**

Independientemente de las fuentes de información de datos meteorológicas se debe considerar que éstos sean de calidad y ello ocurre cuando responden a determinadas necesidades tanto implícitas como explícitas y a las aplicaciones para las que se vaya a utilizar la información, para ello se debe asegurar que los datos cumplan condiciones necesarias (en cuanto a margen de incertidumbre, resolución, continuidad, homogeneidad, representatividad, puntualidad, formato, etc.), su calidad debe ser conocida (dentro de un intervalo) y demostrable, para ello la Organización Mundial de Meteorología esta basado 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 sola estación como procedimientos para validar datos de diferentes variables meteorológicos son cinco: rangos o intervalos, consistencia temporal, persistencia, consistencia interna y consistencia espacial (Estévez, 2008).

El tests de rangos o intervalos comprueba si una observación está dentro de un rango predeterminado que puede ser fijo o dinámico, tal como se muestra en la Tabla 01.

**Tabla 1***Límites físicos de diferentes variables meteorológicas.*

Variable	Unidad	Rango
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 son procedimientos comprueban si la diferencia entre medidas meteorológicas sucesivas excede un valor determinado, en cuyo caso habría que sospechar de ambas medidas, la OMM recomienda una serie de tolerancias propuestas solamente 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 2***Tolerancias propuestas en función del tiempo transcurrido entre medidas consecutivas*

	dt=1 hora	dt=2 horas	dt=3 horas	dt=6 horas	dt=12 horas
<b>T TOL</b>	4 °C	7 °C	9 °C	15 °C	25 °C
<b>T<sub>dew</sub> 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 son procedimientos basados en la consistencia temporal. En lugar de evaluar el excesivo cambio o salto entre observaciones sucesivas, ahora se comprueba la escasa o nula variabilidad de dichos registros. La consistencia interna por su lado son procedimientos que están basados en la verificación de la coherencia física o climatológica de cada variable observada o

también de la consistencia entre variables. La consistencia espacial son algoritmos que sirven para detectar posibles errores comparando los datos de una estación con los valores correspondientes a estaciones vecinas (World Meteorological Organization, 1993).

### **2.3.2. 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 del aire y se deposita en el suelo debido a la acción de la gravedad y comprende la lluvia, el granizo, la nieve, el rocío, la cencellada blanca, la escarcha y la precipitación de la niebla. El instrumento que se utiliza para medir la precipitación caída en un lugar y durante un tiempo determinado se denomina pluviómetro. La precipitación se mide en milímetros de agua o litros caídos por unidad de superficie  $l/m^2$  (volumen/área) o en  $kg/m^2$  (masa/área) para la precipitación líquida (World Meteorological Organization, 2017).

La Humedad es un término utilizado generalmente para enunciar la cantidad de vapor de agua presente en el aire. Existen distintas magnitudes en la medición de la humedad. Razón de mezcla, que consiste en 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, que representa la razón, expresada en porcentaje, entre la presión de vapor observada y la tensión del vapor saturante con respecto al agua a la misma temperatura y presión (WMO, 2015).

El viento es el movimiento del aire en una dirección y una velocidad determinada. La Dirección del viento indica de donde proviene o sopla el viento y es el valor obtenido del promedio de 10

minutos de la dirección del viento, su unidad de medición se representa en grados Dextrorsum (giro en sentido de las manecillas del reloj) donde  $0^\circ$  es norte verdadero, 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 el promedio aritmético de las velocidades medidas en un lapso de 10 minutos, su unidad de medición para estaciones meteorológicas automáticas (EMA) es en km/h., para Estaciones Sinópticas Meteorológicas (ESIME's) es 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, Valenzuela, Marcos, & Gradilla, 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, cuanto se produce mayor agitación, mayor será la temperatura. El 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 cada minuto), su unidad de medición depende de la escala que se elige y puede ser: Escala Celsius ( $^\circ\text{C}$ ) la más utilizada, Escala 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 que contiene el aire. Existen diversas maneras de referirnos al contenido de humedad en la atmósfera como son: la humedad absoluta, la humedad específica y la razón de mezcla. Sin embargo, la medida de humedad que más se utiliza es la denominada humedad relativa, que se expresa en tanto por ciento (%). La humedad se suele medir mediante un

instrumento denominado psicrómetro o higrómetro y se calcula según la siguiente 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 ejerce el aire que nos rodea sobre todos los cuerpos debida a la acción de la gravedad por unidad de superficie, depende de muchas variables, sobre todo de la altitud. El instrumento que se utiliza para medir la presión atmosférica es el barómetro. Tiene diferentes unidades, las más comunes son: atmósferas, mm de mercurio, pascuales, hectopascuales y milibares. La conversión entre unas y otras puede realizarse teniendo en cuenta que: 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 (Hectopascal) (Fundación Española para la Ciencia y la Tecnología, 2004).

La radiación solar es la energía transferida por el Sol a la Tierra que viaja a través del espacio en forma de ondas clasificadas como espectro electromagnético, Las ondas más energéticas son las correspondientes al rango del ultravioleta, seguidas por la luz visible, infrarroja y 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 denominado piranómetro y para registrar la duración e intensidad se usa el heliógrafo. Su unidad de 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 3, se muestra un resumen de los instrumentos de medida y su correspondiente variable meteorológica a medir y las unidades de medida respectiva.

**Tabla 3**  
*Instrumentos de variables meteorológicas*

Instrumento de medida	Variable meteorológica	Unidades de medida
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>

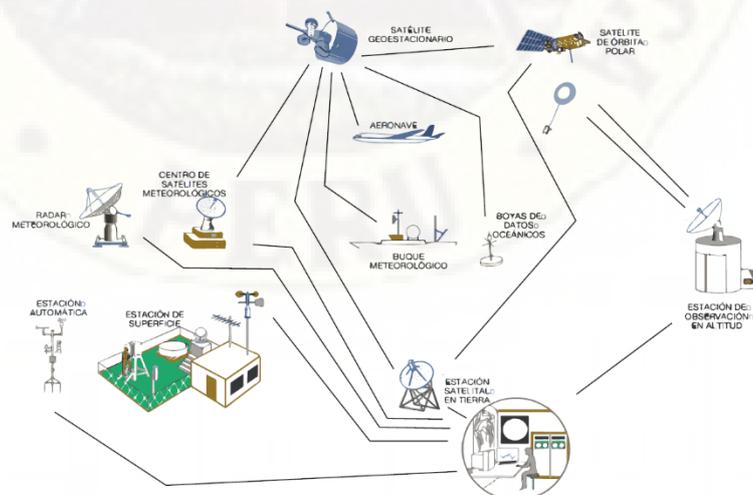
**b) Sistemas de observación de variables meteorológicas**

Para el estudio y medición de variables meteorológicas existen variados sistemas de observación para la medición de datos en función de las consideraciones temporales y espaciales deseadas. El uso de estos sistemas e instrumentos de observación permiten la obtención de indicadores de dichas variables desde distintas perspectivas y con errores distintos, permitiendo tener, de manera global, un mejor entendimiento del comportamiento de las variables meteorológicas. Los satélites meteorológicos son los sistemas de observación más complejos y por la posición que ocupan los satélites, en órbita alrededor de la Tierra, les permite tener una visión

privilegiada, mucho más amplia y de conjunto que la de cualquier aparato situado en la superficie terrestre (Fundación Española para la Ciencia y la Tecnología, 2004).

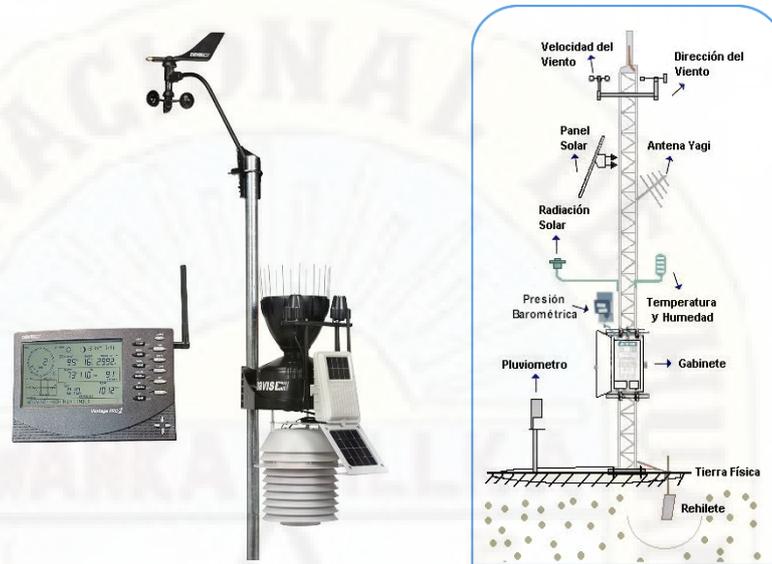
Uno de los sistemas de captura de datos más difundido son las Estaciones Meteorológicas Automáticas – EMA (*Automatic Weather Station - AWS*), que son sistemas basados en radar o detecciones desde satélites en la que se realizan y se transmiten observaciones automáticamente, tal como se muestra en la 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, el conjunto de dispositivos eléctricos que realizan mediciones de las variables meteorológicas de manera automática se denomina Estación Sinóptica Meteorológica (ESIME) tal como se muestra en la Figura 2 (Mendoza, Valenzuela, Marcos, & Gradilla, 2020) (Guerra, 2017).

**Figura 1** Sistema de observación meteorológica



**Figura 2**

*Estación Sinóptica Meteorológica (ESIME)*



**c) Patrones de comportamiento meteorológico**

La naturaleza se comporta a partir de ciclos, tenemos el ciclo del agua o el de la vida, los ciclos del clima también son muy conocidos, alternando edades de hielo con periodos cálidos interglaciares, respecta a la meteorología ocurre algo parecido, existe patrones de comportamiento. La meteorología se estudian la atmósfera desde varias perspectivas. Por una parte, se describen las condiciones del tiempo atmosférico en un espacio y tiempo concreto. Por otro, se desarrolla investigaciones sobre el comportamiento de cada uno de las variables meteorológicas (temperatura, presión, humedad, radiación solar ...) con el fin de descubrir las leyes y patrones respecto a su influencia sobre otras variables y realizar una previsión del tiempo acertada.

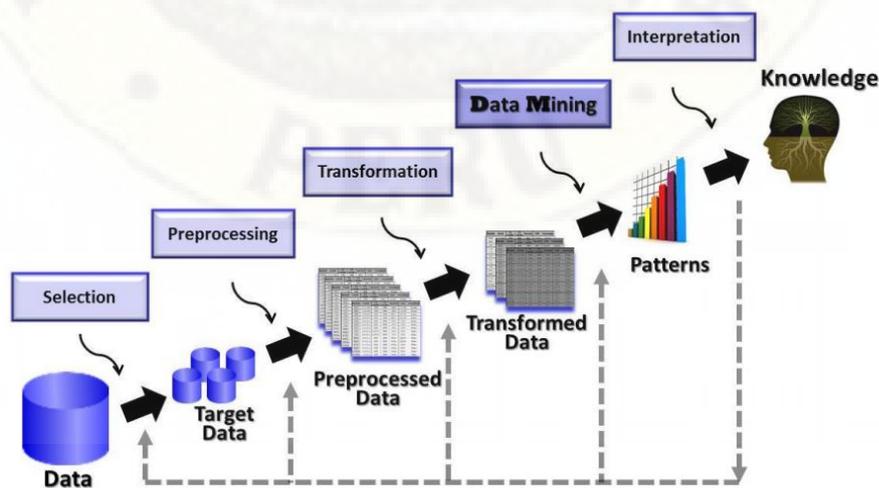
Los fenómenos y alteraciones meteorológicos influyen diariamente en la sociedad y el ambiente, provocando efectos adversos, como beneficiosos; de ahí, la importancia de estudiar los datos meteorológicos a fin de descubrir o conocer su comportamiento, con el objetivo de caracterizar en determinados periodos 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 3 (Zamora, 2018).

**Figura 3**

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



**Fuente:** 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, Bedasa, & Terefa, 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, Muqem, & Javed, 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).

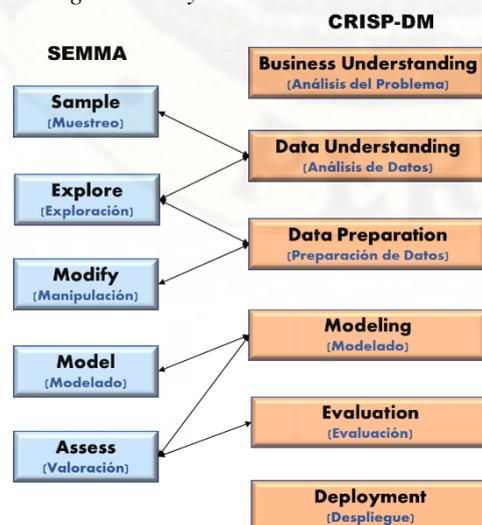
Técnicas de Data Mining fueron utilizados en la zona de Cujae – Cuba con el objetivo de obtener un modelo de datos para estimar el comportamiento de parámetros meteorológicos secundarios a partir de datos de superficie mediante la metodología CRISP-DM y la herramienta de análisis de datos WEKA en la que se utilizó tareas de selección de atributos y de regresión y la técnica de redes neuronales de tipo perceptrón multicapas (Cruz, Zerquera, Morales, & Rosete, 2012).

Para identificar y extraer el comportamiento o generar patrones en torno a datos climatológicos, como la precipitación registrada por el Servicio Meteorológico Nacional (SMN) en los estados de Chiapas, Oaxaca, Tabasco y Veracruz se utilizó la herramienta computacional 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 metodología para emprender proyectos para procesar datos usando 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 4, se muestra las fases e interrelaciones de éstas dos metodologías (Montequín, 2002) (Zamora, 2018).

**Figura 4**

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



**Fuente:** 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 y que están organizados en función al tipo de aprendizaje que utilizan y que puede ser supervisado y no supervisado. Los algoritmos de aprendizaje supervisado basan su aprendizaje en un conjunto de datos de entrenamiento previamente etiquetado en variables en la que se conoce el valor del atributo objetivo, a predecir o variable dependiente, asocian entradas con salidas, estos algoritmos a su vez pueden ser de regresión o de clasificación. Los algoritmos de aprendizaje no supervisados basan su proceso de entrenamiento en un conjunto de datos sin etiquetas, variables o clases previamente definidas, su función es el agrupamiento, segmentación (clúster) para encontrar grupos similares en el conjunto de datos que pueden ser jerárquicos o particionales (García, 2020).

**g) Análisis de series temporales**

Una serie de tiempo, 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, Bedasa, & Terefa, 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).

## **2.4. Marco filosófico**

### **2.4.1. Encuadramiento Epistemológico del Problema**

Planteado el problema de investigación, así como los objetivos generales y específicos para su solución, a continuación se realiza un encuadramiento epistemológico del problema para su solución, intentando determinar por las características del problema, a qué tipo de criterio o tendencia se ubica para su explicación y solución, por tanto se expondrá de manera concisa, a continuación las diferentes corrientes filosóficas de modo que se pueda explicar y justificar el problema de investigación planteado.

### **2.4.2. El Pragmatismo**

El fundador del pragmatismo se considera al filósofo norteamericano William James († 1910). El pragmatismo abandona el concepto de la verdad en el sentido de la concordancia entre el pensamiento y el ser. Pero el pragmatismo no se detiene en esta negación, sino que reemplaza el concepto

abandonado por un nuevo concepto de la verdad. Según él, verdadero significa útil, valioso, fomentador de la vida. El pragmatismo modifica de esta forma el concepto de la verdad, porque parte de una determinada concepción del ser humano. Según él, el hombre no es en primer término un ser teórico o pensante, sino un ser práctico, un ser de voluntad y acción. El error fundamental del pragmatismo consiste en no ver la esfera lógica, en desconocer el valor propio, la autonomía del pensamiento humano (Hessen, 2006).

#### **2.4.3. Racionalismo**

La forma más antigua del racionalismo se encuentra en Platón. El Racionalismo es la posición epistemológica que ve en el pensamiento, en la razón, la fuente principal del conocimiento humano, se llama racionalismo (de ratio = razón). Según él, un conocimiento sólo merece, en realidad, este nombre cuando es lógicamente necesario y universalmente válido. Cuando nuestra razón juzga que una cosa tiene que ser así y que no puede ser de otro modo; que tiene que ser así, por tanto, siempre y en todas partes, entonces y sólo entonces nos encontramos ante un verdadero conocimiento, en opinión del racionalismo. El mérito del racionalismo consiste en haber visto y subrayado con energía la significación del factor racional en el conocimiento humano. Un defecto del racionalismo consiste en respirar el espíritu del dogmatismo. Cree poder penetrar en la esfera metafísica por el camino del pensamiento puramente conceptual (Hessen, 2006).

#### **2.4.4. El Empirismo**

Su fundador es John Locke (1632-1704). El empirismo opone a la tesis del racionalismo (según la cual el pensamiento, la razón, es la verdadera fuente del conocimiento) la antítesis que dice: la única fuente del conocimiento humano es la experiencia. En opinión del empirismo, no hay ningún patrimonio a priori de la razón. La conciencia cognoscente no saca sus contenidos 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, proceden de la experiencia (Hessen, 2006).

#### **2.4.5. Intelectualismo:**

El fundador del Intelectualismo es Aristóteles. El intelectualismo es aquella dirección epistemológica que realiza uno de estos intentos de mediación entre el racionalismo y el empirismo. Mientras el racionalismo considera el pensamiento como la fuente y la base del conocimiento y el empirismo la experiencia, el intelectualismo es de opinión que ambos factores tienen parte en la producción del conocimiento. El intelectualismo sostiene con el racionalismo que hay juicios lógicamente necesarios y universalmente válidos, y no sólo sobre los objetos ideales, (esto lo admiten también los principales representantes del empirismo), sino también sobre los objetos reales. Pero mientras que el racionalismo consideraba los elementos de estos juicios, los conceptos, como un patrimonio a priori de nuestra razón, el intelectualismo lo deriva de la experiencia (Hessen, 2006).

#### **2.4.6. 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 del conocimiento. Pero el apriorismo define la relación entre la experiencia y el pensamiento en un sentido directamente opuesto al intelectualismo. Como ya dice el nombre de apriorismo, nuestro conocimiento presenta, en sentir de esta dirección, elementos a priori, independientes de la experiencia. Esta era también la opinión del racionalismo. Pero mientras éste consideraba los factores a priori como contenidos, como conceptos perfectos, para el apriorismo estos factores son de naturaleza formal. No son contenidos sino formas del conocimiento. Estas formas reciben su contenido de la experiencia, y en esto el apriorismo se separa del racionalismo y se acerca al empirismo. Los factores a priori semejan en cierto sentido recipientes vacíos, que la experiencia 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).

#### **2.4.7. El Objetivismo**

Platón es el primero que ha defendido el objetivismo. Según esta corriente filosófica, el objeto es el decisivo entre los dos miembros de la relación cognoscitiva. El objeto determina al sujeto. Este ha de regirse por aquél. El sujeto toma sobre sí en cierto modo las propiedades del objeto, las reproduce. Esto supone que el objeto hace frente como algo acabado, algo definido de suyo, a la conciencia cognoscente. Justamente en esto reside la idea central del objetivismo. Según él, los objetos son algo dado, algo que presenta una estructura totalmente definida, estructura que es reconstruida, digámoslo así, por la conciencia cognoscente (Hessen, 2006).

#### **2.4.8. El Realismo**

El Realismo aquella posición epistemológica según la cual hay cosas reales, independientes de la conciencia. Esta posición admite diversas modalidades. La primitiva, tanto histórica como psicológicamente, es el realismo ingenuo. Este realismo no se encuentra influido aún por ninguna reflexión crítica acerca del conocimiento. El problema del sujeto y el objeto no existe aún para él. No distingue en absoluto entre la percepción, que es un 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 como identifica los contenidos de la percepción con los objetos, atribuye a éstos todas las propiedades encerradas en aquéllos. Las cosas son, según él, exactamente tales como las percibimos. Los colores que vemos en ellas les pertenecen como cualidades objetivas. Lo mismo pasa con su sabor y olor, su dureza o blandura, etcétera (Hessen, 2006).

#### **2.4.9. El Fenomenalismo**

La teoría del fenomenalismo, ha sido desarrollada por Kant. El fenomenalismo es la teoría según la cual no conocemos las cosas como son en sí, sino como nos aparecen. Para el fenomenalismo hay cosas reales, pero no podemos conocer su esencia. Sólo podemos saber "que" las cosas son, pero no "lo que" son. El fenomenalismo coincide con el realismo en admitir cosas reales; pero coincide con el idealismo en limitar el conocimiento a la conciencia, al mundo de la apariencia, de lo cual resulta inmediatamente la incognoscibilidad de las cosas en sí (Hessen, 2006).

#### **2.4.10. ¿Qué es un Metalenguaje?**

El metalenguaje, según la Real Academia Española, se define como el lenguaje que se usa para hablar del lenguaje, es decir, es el lenguaje utilizado para hablar de los aspectos propios de una lengua o para describirla en su conjunto (ASALE, s. f.).

Todo lenguaje tiene la capacidad de tomar como referencia o directriz un objeto. De tal forma que, todo lenguaje que se centre en el estudio de otro lenguaje, se considera un metalenguaje. Por tanto, se puede decir, que el metalenguaje, se basa en la conciencia adquirida sobre el lenguaje que se usa en cualquier tipo de actividad. Su función es la concienciación de sus componentes para adquirir dominio sobre ellos. El metalenguaje, tal y como señala (Tunmer, Pratt, Herriman, & Bowey, 1984), es uno de los componentes de la metacognición

#### **2.4.11. Sistema Procesador de Operaciones Metalingüísticas**

*(Metalinguistic Operation Processor, o MOP)*

El sistema Procesador de Operaciones Metalingüísticas (MOP) de acuerdo a (Penagos, 2005) extrae enunciados metalingüísticos y definiciones de documentos técnicos, utilizando tanto autómatas de estados finitos como algoritmos de aprendizaje automático. Este sistema se

diferencia de otros sistemas terminográficos en que no se basa simplemente en regularidades sintácticas, morfológicas o semánticas implícitas de la formación de los términos, las estructuras semánticas o los conceptos, sino que explota la dimensión más discursiva de actos de habla en los cuales los términos son establecidos, modificados o valorados de manera explícita por los propios participantes en la actividad científica.

Una vez seleccionados los enunciados, el sistema realiza un procesamiento de base lingüística para crear bases semiestructuradas de información terminológica llamadas Bases de Información Metalingüística (Metalinguistic Information Databases, o MIDs)

## **2.5. Formulación de hipótesis**

### **2.5.1. Hipótesis General**

Los patrones de comportamiento a partir de datos de variables meteorológicas pueden ser obtenidos mediante técnicas de Data Mining en la ciudad de Huancavelica, 2018 – 2019.

### **2.5.2. Hipótesis Específicas**

- a) Los clústeres formados mediante técnicas de Data Mining, a partir de valores meteorológicos diarios en la ciudad de Huancavelica, 2018–2019, presentan diferencias significativas
- b) Las variables meteorológicas procesados mediante técnicas de Data Mining y los diferentes meses del año en la ciudad de Huancavelica, 2018 – 2019, presentan diferencia significativa.

## **2.6. Identificación de variables**

### **2.6.1. Variable1**

Variables meteorológicas en la ciudad de Huancavelica 2018–2019.

## 2.7. Definición operativa de variables e indicadores

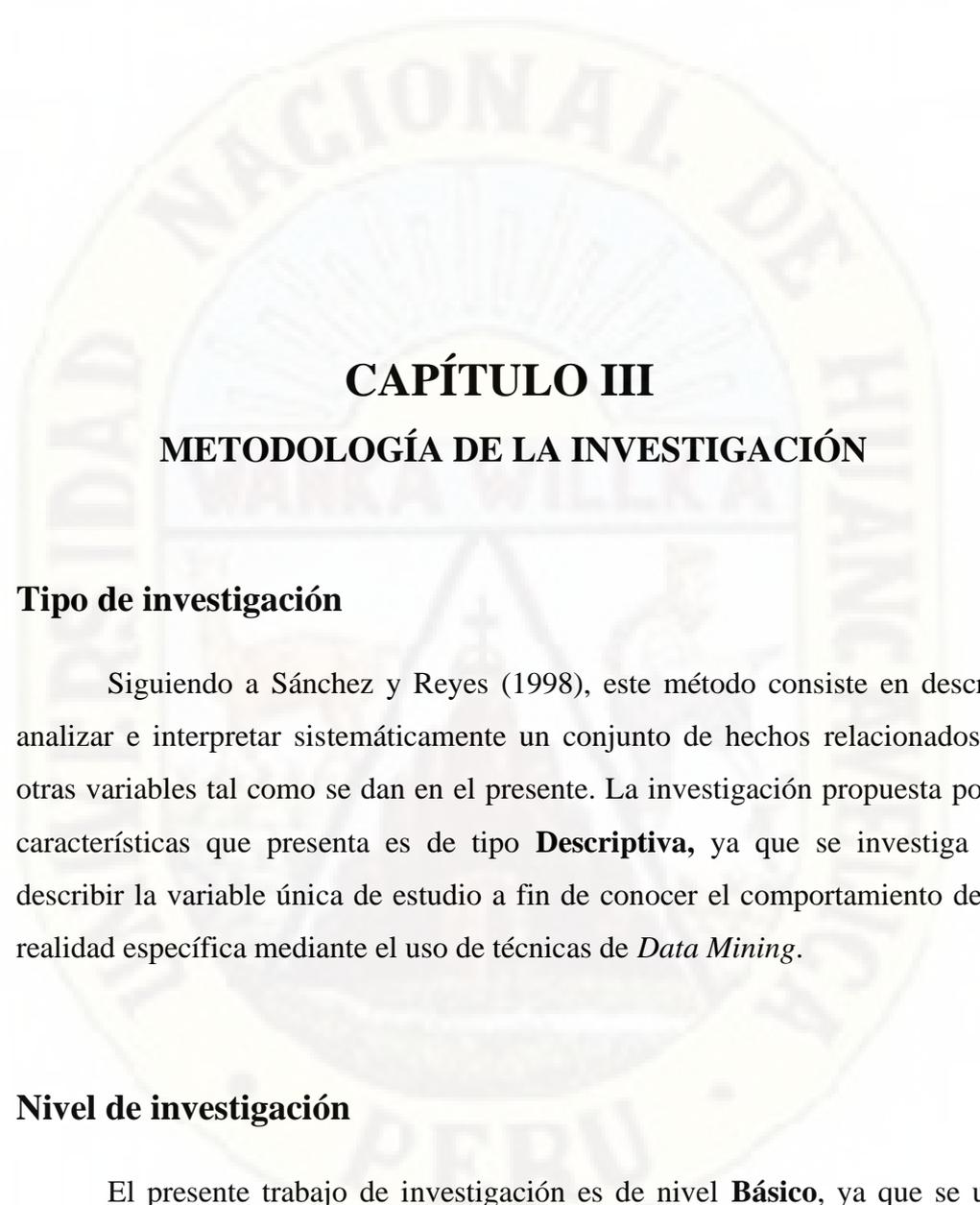
### 2.7.1. Variable I:

Variables meteorológicas en la ciudad de Huancavelica 2018–2019 (Temperatura Ambiental, Precipitación Pluvial, Humedad relativa, Presión atmosférica, Radiación Solar, Radiación Ultravioleta y Velocidad del viento)

**Tabla 4**

*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	Estación	mm
	Humedad relativa	Humedad porcentual promedio en 24Hr.	Meteorológica Automatizada Weather Link	%
	Presión atmosférica	Presión atmosférica promedio de 24Hr.	Marca DAVIS Modelo Vantage Pro	Mm de Hg
	Radiación Solar	Radiación acumulada en 24Hr.	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



## **CAPÍTULO III**

### **METODOLOGÍA DE LA INVESTIGACIÓN**

#### **3.1. Tipo de investigación**

Siguiendo a Sánchez y Reyes (1998), este método consiste en describir, analizar e interpretar sistemáticamente un conjunto de hechos relacionados con otras variables tal como se dan en el presente. La investigación propuesta por las características que presenta es de tipo **Descriptiva**, ya que se investiga para describir la variable única de estudio a fin de conocer el comportamiento de una realidad específica mediante el uso de técnicas de *Data Mining*.

#### **3.2. Nivel de investigación**

El presente trabajo de investigación es de nivel **Básico**, ya que se usará técnicas de *Data Mining* para extraer patrones de comportamiento de datos meteorológicos para que a posteriori se pueda realizar su respectiva gestión ambiental y toma de decisiones en aspectos que dependen del comportamiento de las variables meteorológicas en la ciudad de Huancavelica.

### 3.3. Métodos de investigación

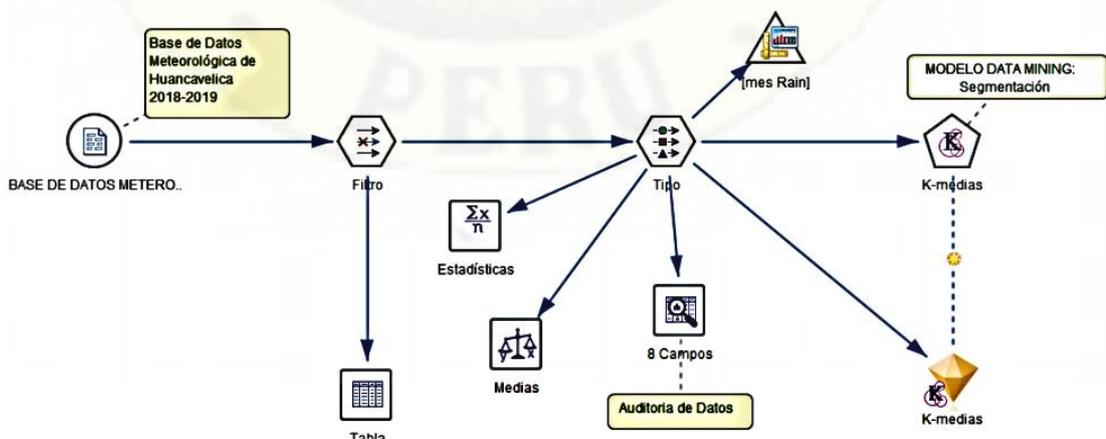
El método a usarse en la presente investigación será el método **Científico**, de manera particular el método longitudinal y series de tiempo ya que se usará datos meteorológicos históricos a fin de extraer patrones de comportamiento de datos meteorológicos en la ciudad de Huancavelica, para su posterior uso en la toma de decisiones y una adecuada gestión ambiental. Los datos de las variables meteorológicas serán manipulados para descubrir patrones mediante técnicas de *Data Mining* y describir comportamientos ocultos de la variable en estudio.

### 3.4. Diseño de investigación

En la presente investigación se usó un diseño modelado mediante el software SPSS Modeler, que inicia desde la base de datos recolectados por la estación meteorológica automática, los filtros usados de las bases de datos, el pre proceso de datos, el modelo algorítmico usado (K-Means) hasta el descubrimiento de patrones planteadas en la investigación, la misma que se muestra en la siguiente Figura 5.

**Figura 5**

*Diseño de la investigación mediante SPSS Modeler para determinar Patrones Meteorológicos en la ciudad de Huancavelica 2018-2019*



## **3.5. Población, muestra y muestreo**

### **3.5.1. Población**

Teniendo en cuenta que la población o universo de estudio de una investigación es el conjunto de individuos que comparten al menos una característica común, la población objetivo que se tuvo referencia para el presente estudio estuvo conformado por los diferentes momentos horarios que se evaluó las variables meteorológica en ciudad de Huancavelica en el punto donde se encuentra ubicada la estación meteorológica automática que permitió registrar los datos meteorológicos para su posterior análisis mediante técnicas de Data Mining y descubrir los posibles patrones que estos presentan.

### **3.5.2. Muestra**

A partir de la población establecida se realizó la toma de datos de la muestra, en momentos horarios en las que se evaluó las variables meteorológica que estuvo conformada por el periodo de estudio muestral de un año establecido del primero de agosto del 2018 al 31 de julio del 2019 en la que se registró los datos usando una estación meteorológica automática instalada en la ciudad de Huancavelica ubicada en la ciudad universitaria de Paturpampa, en la parte superior del laboratorio central, distrito de santa Ana, provincia de Huancavelica y departamento de Huancavelica .

### **3.5.3. Muestreo**

La técnica de muestreo del presente trabajo fue el muestreo aleatorio sistémico, ya que la estación meteorológica estuvo programada para registrar datos meteorológicos de a ser procesados sistemáticamente cada 60 min. durante 12 meses y 358 días del año desde el primero de agosto del 2018 al 31 de julio del 2019.

## **3.6. Técnicas e instrumentos de recolección de datos**

### **3.6.1. Técnicas**

Las técnicas usadas en la presente investigación fueron la observación, para ello se usó instrumentos implementados en una estación meteorológica de tipo automática, las mismas que se encuentran validados por el Servicio Nacional de Meteorología e Hidrología del Perú - SENAMHi.

### **3.6.2. Los instrumentos**

Los instrumentos y equipos para la medición de las diferentes variables meteorológicas que pueden ser primero las manuales o convencionales y las segundas automáticas, son cada día más precisos, siendo la segunda la de mayor eficiencia, esta característica es la que día a día permite realizar predicciones cada vez más rápidas y con mayor exactitud. El instrumento usado en la presente investigación para la recolección de datos de la variable en estudio fue una estación meteorológica automatizada Weather Link, Marca DAVIS, Modelo Vantage Pro y una consola Vantage Pro, tal como se muestra en la Figura 6, las mismas que se encuentran dotados de sensores para registrar los diferentes parámetros meteorológicos ambientales tales como: la temperatura ambiental (promedio, máxima, mínima e interna) presión atmosférica, humedad atmosférica, velocidad del viento, dirección del viento, radiación solar, radiación ultra violeta y precipitación pluvial, entre otras, todo ello de modo automatizada y digitalizada para su posterior procesamiento y análisis.

**Figura 6**

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



Se usó una estación automatizada en vez de una convencional por tener estas primeras mayor precisión y no requiere de un registrador humano evitando sesgos en la toma de datos, entre otras variables, esta estación meteorológica automatizada proporcionó datos sobre la temperatura ambiental, presión atmosférica, humedad atmosférica, velocidad del viento, radiación solar, radiación ultra violeta y precipitación pluvial, las mismas que fueron registradas con un intervalo de una hora y posteriormente se realiza la descarga de datos de la consola Vantage Pro como texto plano a una microcomputadora personal durante periodos mensuales durante el tiempo de estudio que fue de un año, a una hoja de cálculo electrónico.

### **3.7. Técnicas de procesamiento y análisis de datos**

Dentro de las técnicas de procesamiento y análisis de datos se usó en principio la hoja de cálculo electrónico Microsoft Excel usando como entrada los datos importados de la consola Vantage Pro en texto plano con formato CSV-UTF-8 delimitado por comas, recolectada para el análisis descriptivo y cálculo de los estadígrafos correspondientes, fueron evaluados las medidas de tendencia central

de los datos meteorológicos (la media aritmética, la mediana y la moda), las medidas de Dispersión (La varianza, la desviación estándar y el coeficiente de variabilidad) a fin de describir de manera primaria y convencional los datos, para la descripción de la variable velocidad y dirección del viento y diseño de la rosa de viento anual y mensual se usó WRPLOT View - Lakes Environmental Software – Wind Rose for Meteorological Data, para los 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*), tal como muestra la Figura 7.

**Figura 7**

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

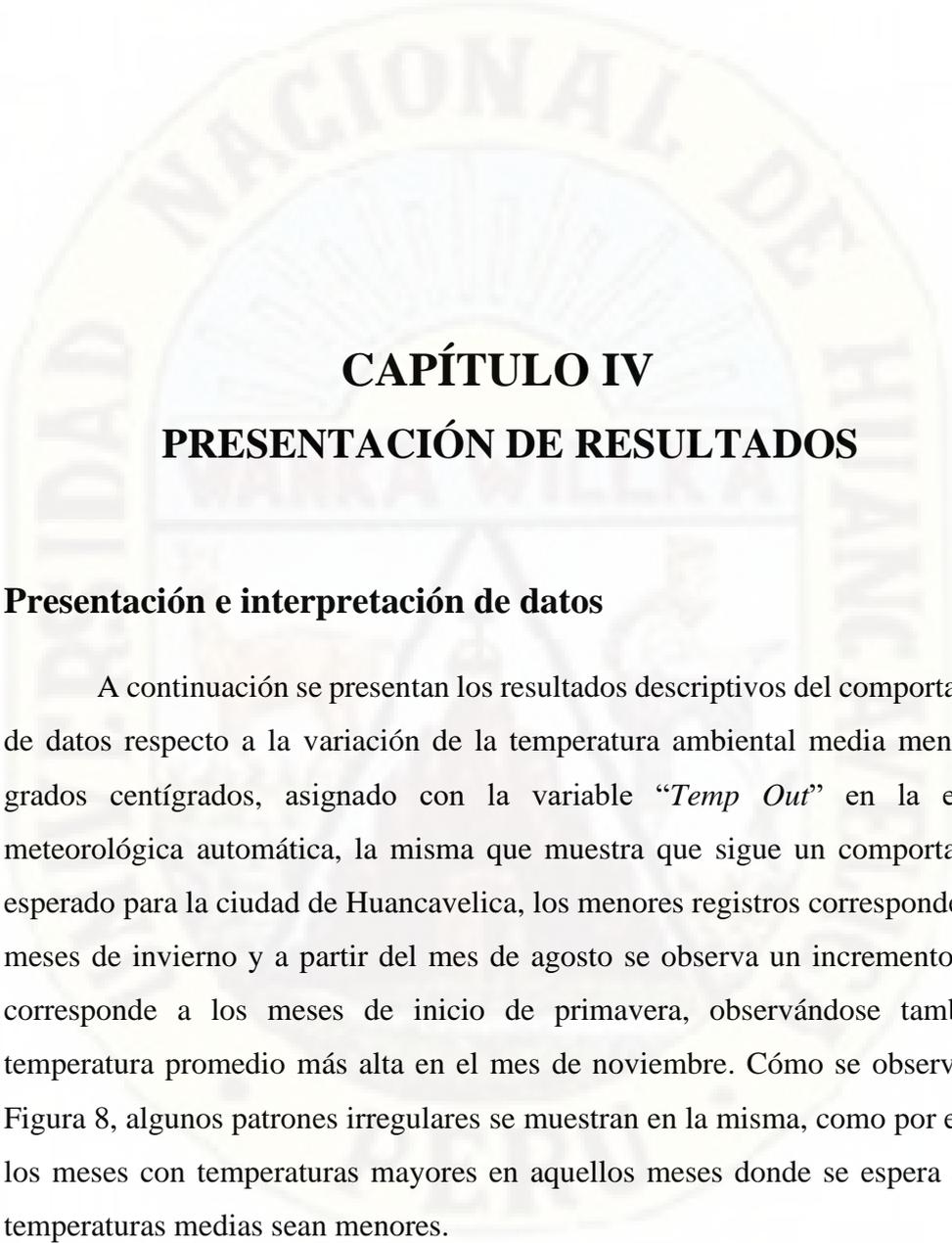


### 3.8. Descripción de la prueba de Hipótesis

Para las pruebas de las hipótesis, se realizó en función al análisis de datos de los resultados de los programas y aplicaciones informáticas usadas, de los reportes gráficos que nos permitieron extraer conocimiento oculto a la luz de la estadística convencional, tales como patrones meteorológicos, asociación de variables meteorológicas, se usó algoritmos de árboles de decisión, técnicas de segmentación mediante el algoritmo de K-means implementado en SPSS modeler, tendencias entre otros análisis de Data Mining.

Mediante SPSS statistics se comprobó la primera hipótesis específica respecto al grado de significancia de los cluster o grupos formados mediante técnicas y algoritmos de Data Mining, luego la segunda hipótesis específica se comprobó mediante SPSS Modeler para determinar la significancia de las variables meteorológicas en los meses del periodo de estudio, que serán mostrados en los resultados de la presente investigación.





## **CAPÍTULO IV**

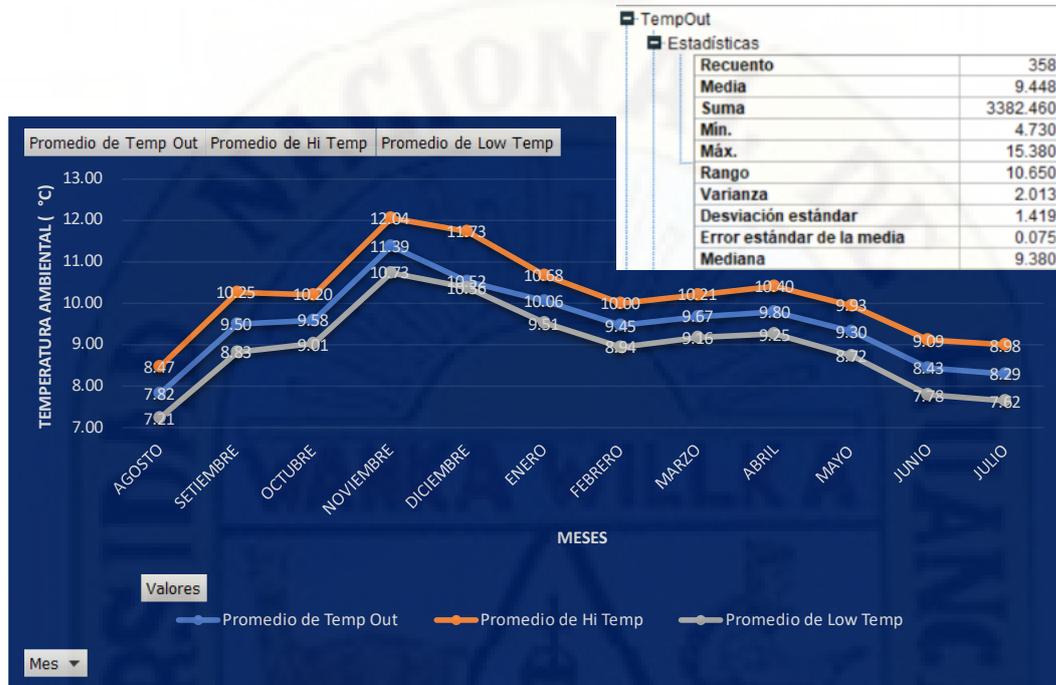
### **PRESENTACIÓN DE RESULTADOS**

#### **4.1. Presentación e interpretación de datos**

A continuación se presentan los resultados descriptivos del comportamiento de datos respecto a la variación de la temperatura ambiental media mensual en grados centígrados, 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. Como 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**

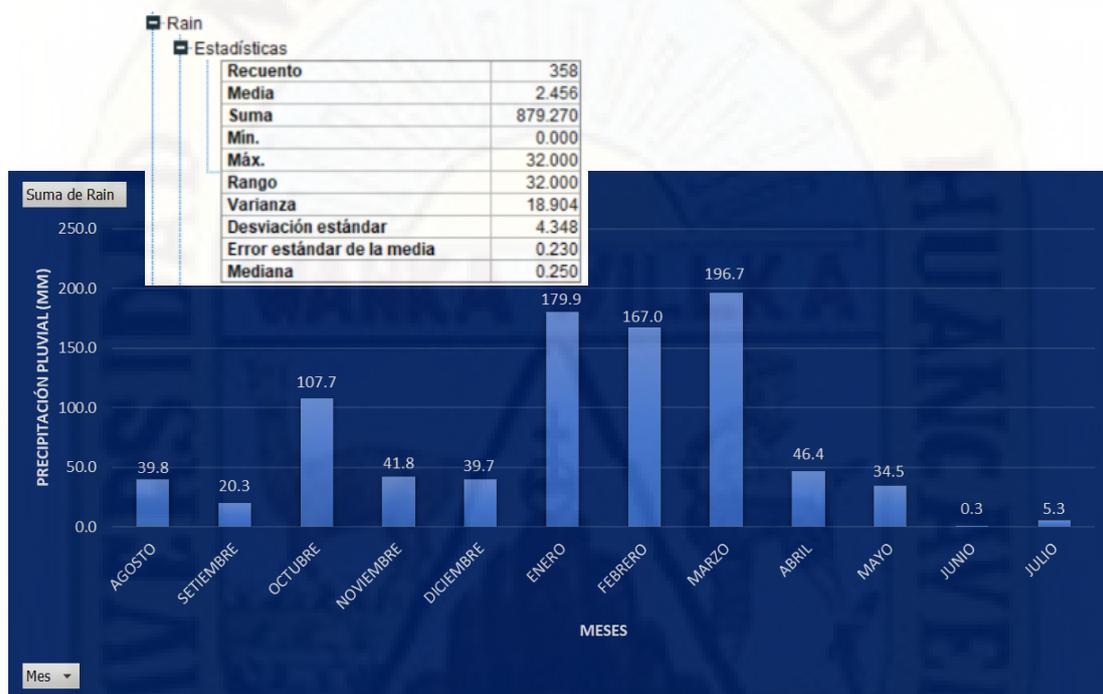
Temperatura Ambiental de la ciudad de Huancavelica, periodo 2018-2019.



En la siguiente Figura 9 se presentan los resultados descriptivos del comportamiento de datos respecto a 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.

**Figura 9**

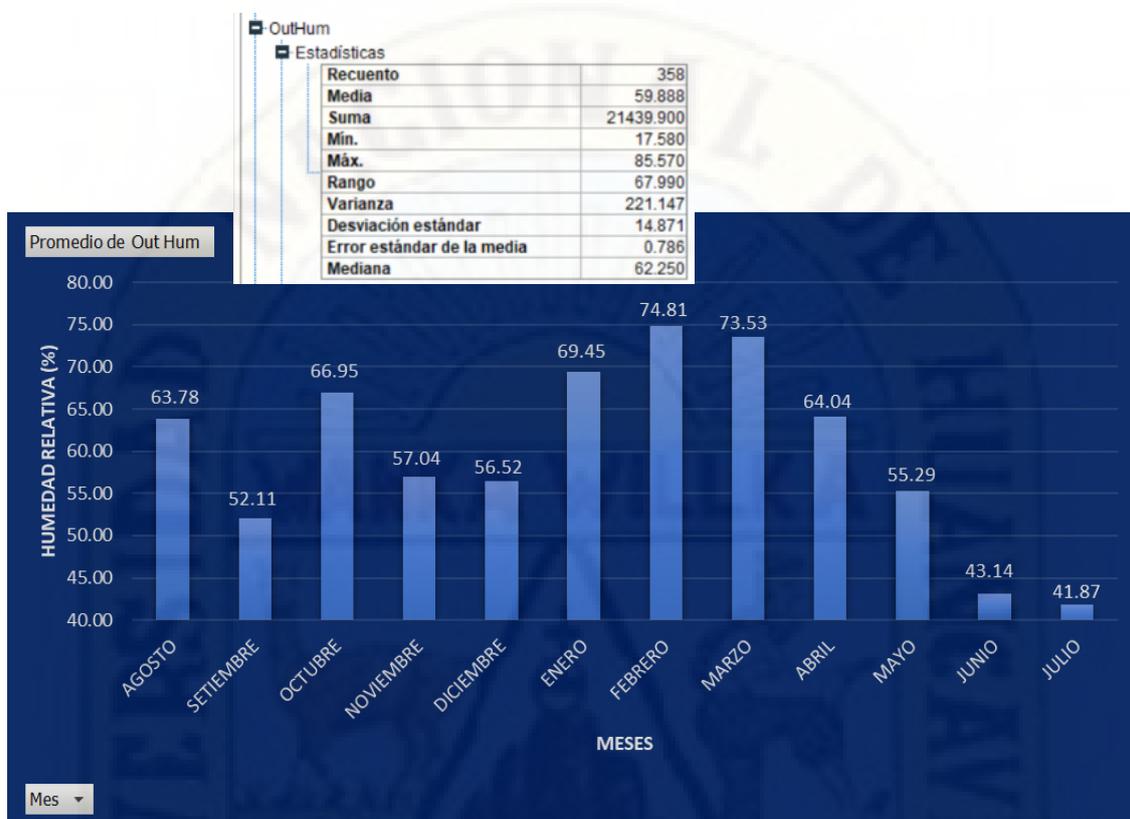
*Precipitación Pluvial de la ciudad de Huancavelica, periodo 2018-2019.*



A continuación se presenta los resultados descriptivos del comportamiento de datos respecto a 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**

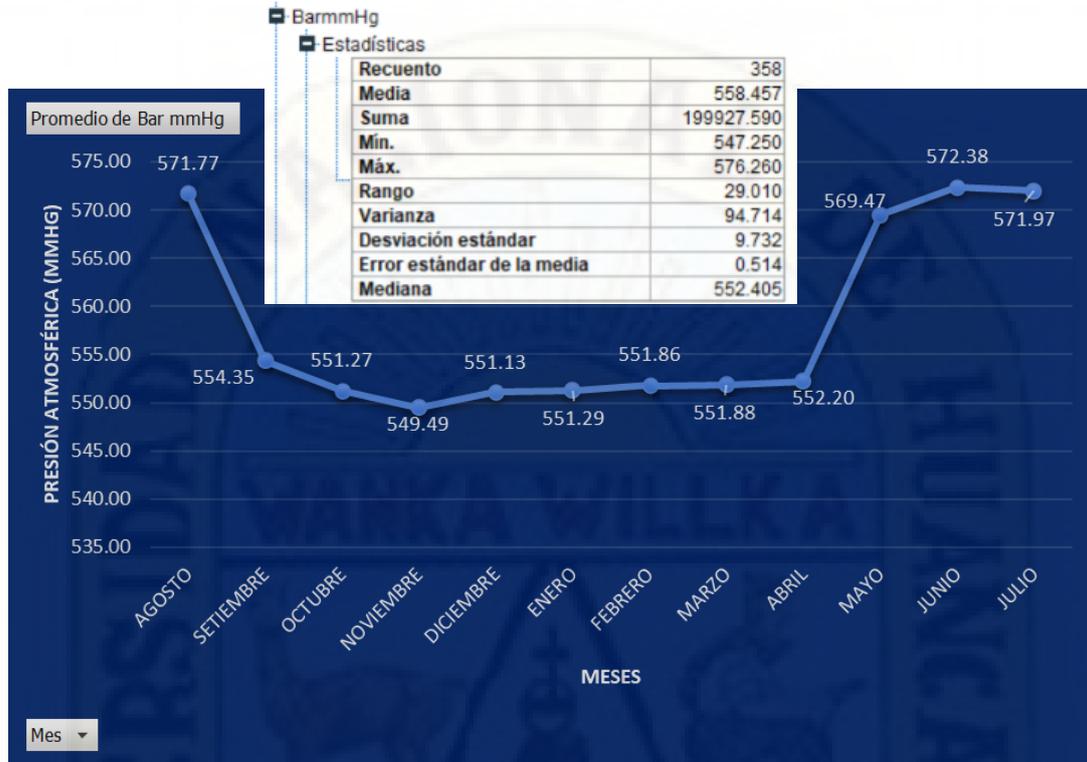
*Humedad relativa de la ciudad de Huancavelica, periodo 2018-2019.*



El siguiente resultado de datos procesados para su análisis descriptivo, muestra 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.*



A continuación, se presenta los resultados procesados de 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  $16.44 \text{ MJ}/m^2$ .

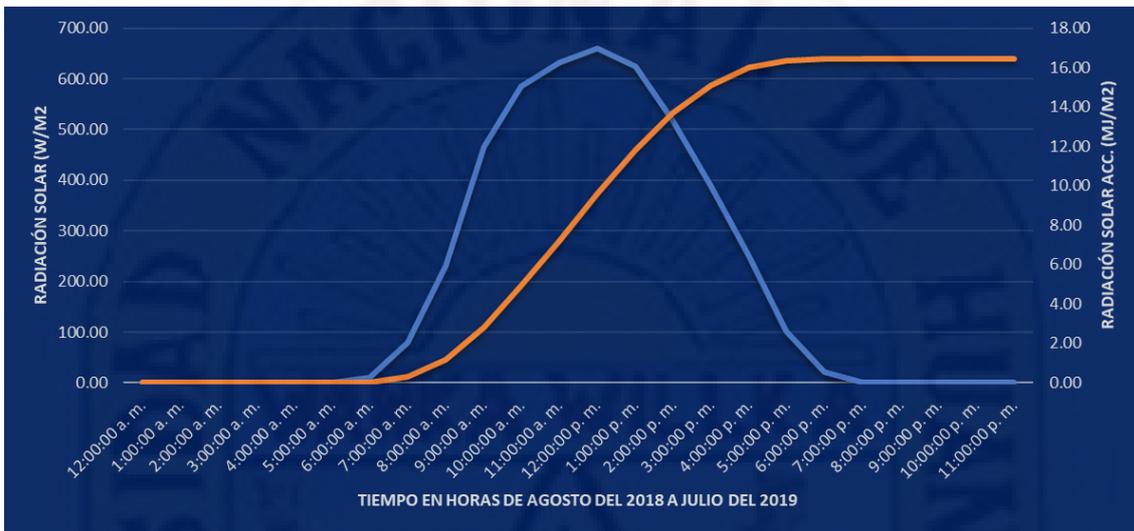
**Figura 13**

*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.*



A continuación, se presentan los resultados de datos procesados respecto al índice de radiación ultravioleta, asignado con la variable “*UV Index*” obtenidos en la estación meteorológica automática, en la que se ha considerado 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. y las 3 p.m.

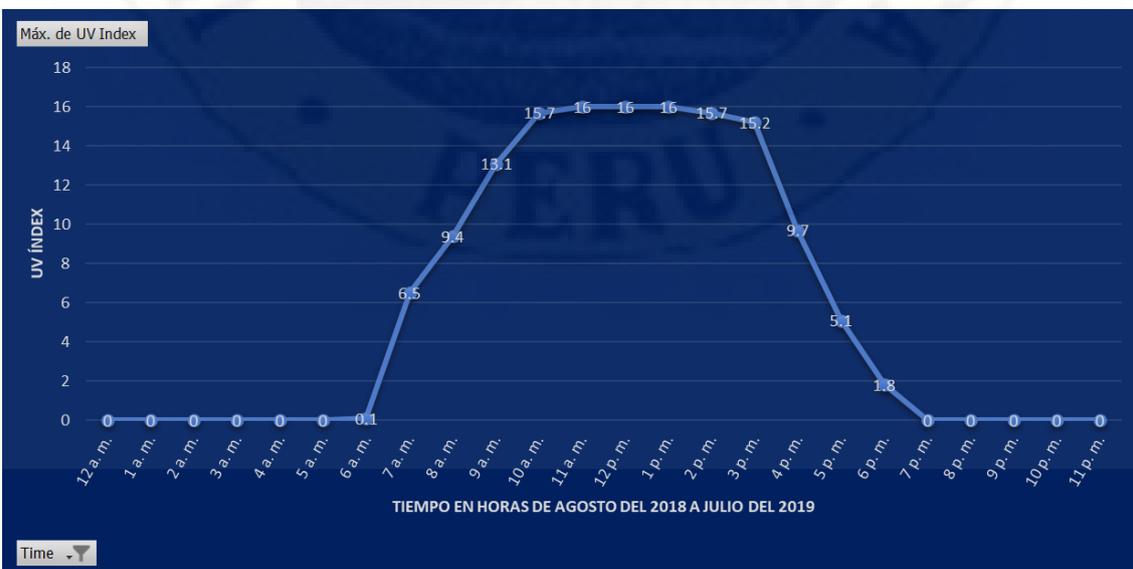
**Figura 15**

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



**Figura 16**

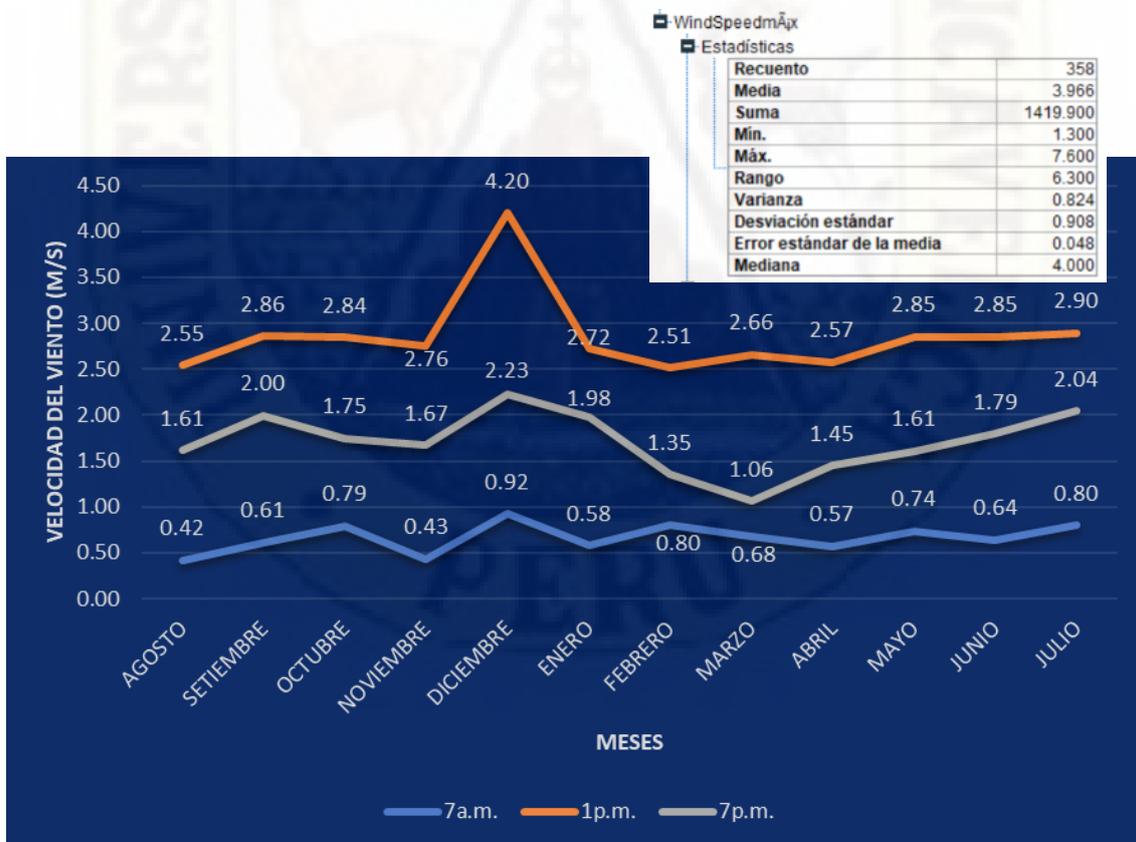
*Radiación Ultravioleta de la ciudad de Huancavelica por horas, periodo 2018-2019.*



A continuación se muestran los resultados de datos procesados de la variable velocidad del viento, asignado con la variable “Wind Speed” obtenidos en la estación meteorológica automática, 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**

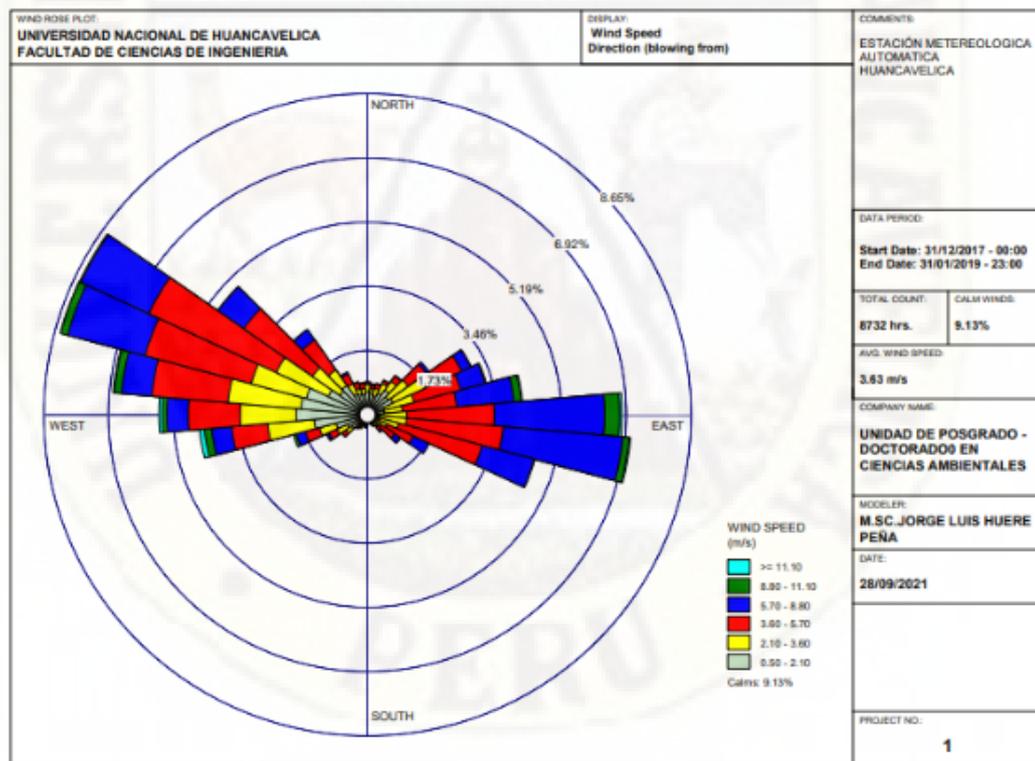
*Velocidad promedio del viento mensual a las 7, 13 y 19 horas en la ciudad de Huancavelica, periodo 2018-2019.*



El Gráficos 18 muestran las Rosas de vientos se puede apreciar el origen del viento (*blowing from*), El Gráficos 19 muestran las Rosas de vientos se puede apreciar la dirección a dónde va el viento (*blowing to*) estas Rosas de Vientos (Wind Rose) fueron diseñadas a partir de la data extraída de la estación meteorológica automática de Huancavelica respecto a los 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**

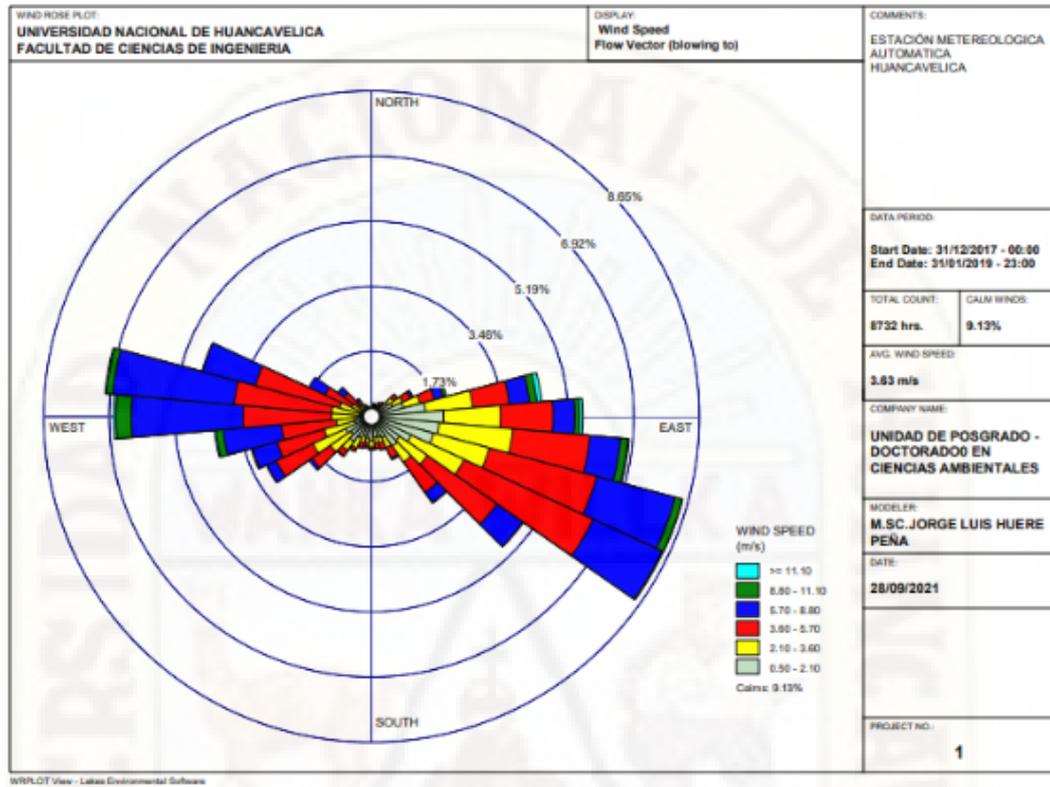
Rosa de viento: “blowing from” en la ciudad de Huancavelica- periodo 2018-2019.



**Fuente:** Elaboración propia con ayuda de WRPLOT lakes environmental software.

**Figura 19**

Rosa de viento: "blowing to" en la ciudad de Huancavelica- periodo 2018-2019.

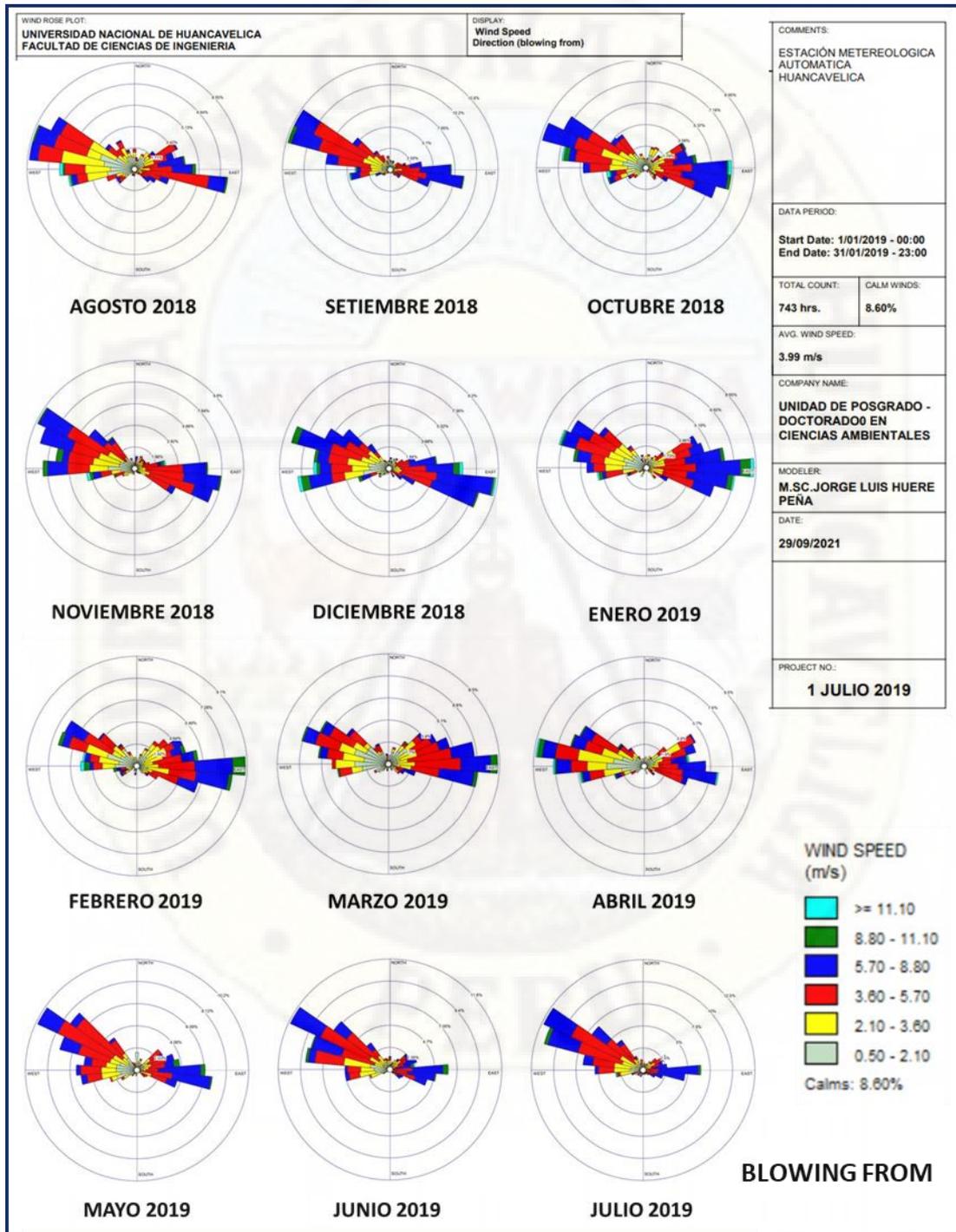


**Fuente:** Elaboración propia con ayuda de WRPLOT lakes environmental software.

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**

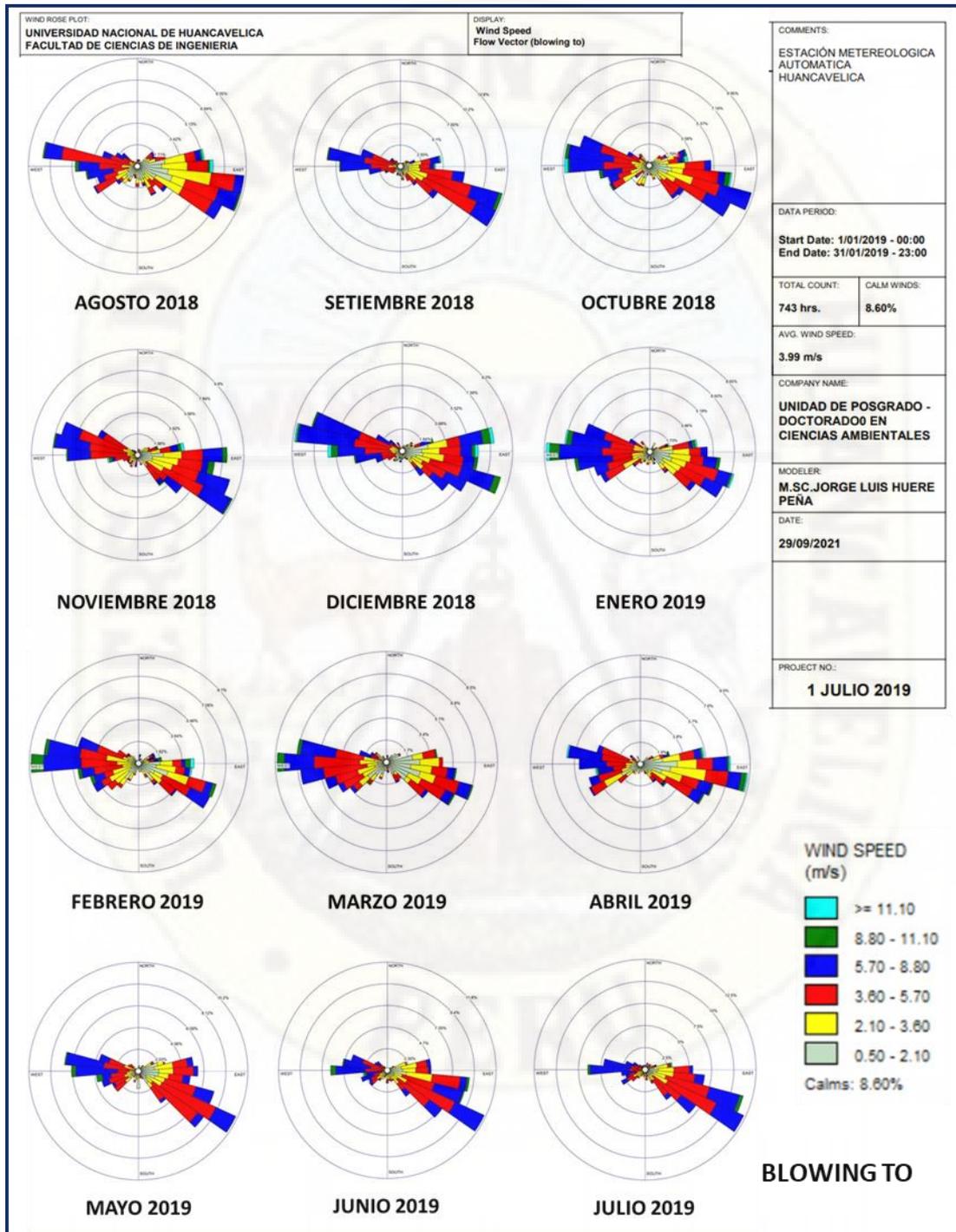
*Rosa de viento: "blowing from" en la ciudad de Huancavelica por meses 2018-2019*



**Fuente:** Elaboración propia con ayuda de WRPLOT Lakes Environmental Software.

**Figura 21**

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

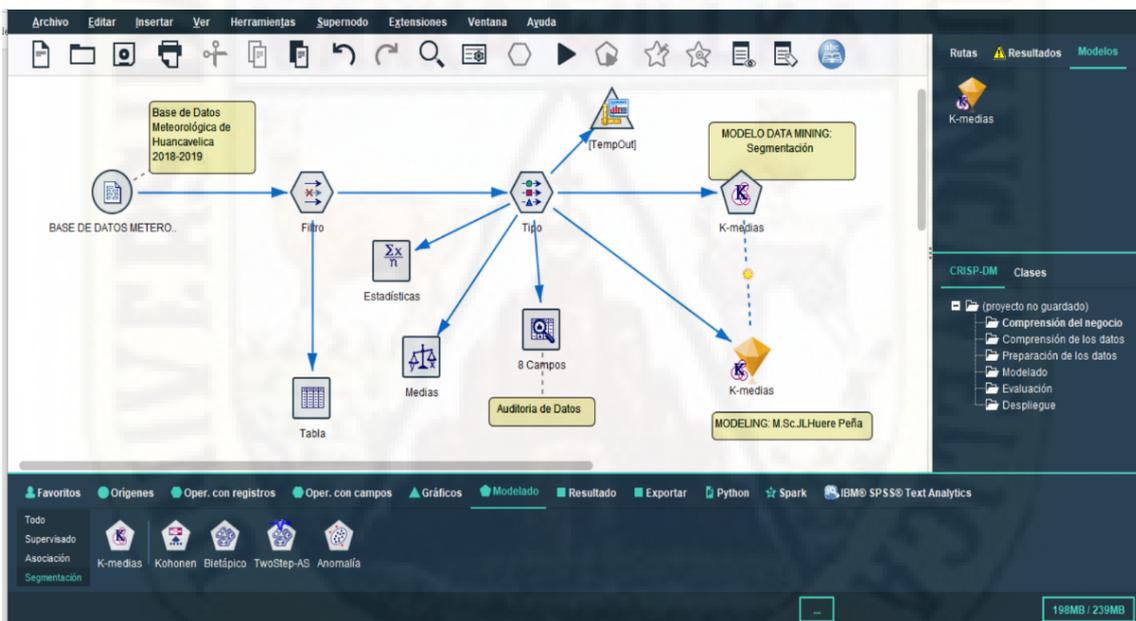


Fuente: Elaboración propia con ayuda de WRPLOT Lakes Environmental Software.

Para realizar el proceso de Data Mining en primer lugar se realizó 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 5***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 6***Historial de iteraciones 04 clúster de las variables meteorológicas*

Iteración	Cambiar en centros de clústeres			
	1	2	3	4
	444,190	877,089	238,500	358,438
	46,868	491,983	589,686	133,729
	16,350	187,032	195,020	101,100
	2,426	37,149	284,133	30,913
	3,522	19,519	193,986	23,344
	8,151	38,734	209,684	23,539
	27,130	18,816	219,190	11,621
	58,126	36,011	212,185	6,979
	68,935	17,329	110,735	21,568
	85,833	,000	154,274	16,781

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 7**  
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 8**  
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 9**  
*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
,000	27,170	170,943	58,609	75,455	5,427	76,018	11,956	137,605	325,768	
,000	34,475	185,357	61,072	64,712	33,980	65,559	,000	80,386	116,958	
,000	23,389	49,326	12,170	30,292	63,649	29,686	,000	36,463	,000	
,000	17,336	110,240	11,794	17,464	14,414	,000	,000	44,947	,000	
,000	17,328	50,593	6,037	11,519	14,923	,000	,000	35,778	,000	
,000	23,992	65,222	5,963	5,368	,000	,000	,000	46,489	,000	
,000	12,030	35,735	,000	,000	,000	,000	,000	26,344	,000	
,000	,000	47,776	,000	,000	,000	,000	,000	20,344	,000	
,000	,000	38,703	,000	,000	,000	,000	,000	21,282	,000	
,000	5,411	12,417	,000	,000	,000	,000	,000	13,539	,000	

- 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 13,532. La iteración actual es 10. La distancia mínima entre los centros iniciales es 708,297.
- b. distancia mínima entre los centros iniciales es 708,297.

**Tabla 10**  
*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 11**

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

Número de casos en cada clúster		Número de casos en cada clúster	
Clúster	114,000	Clúster	1,000
	54,000		60,000
	48,000		24,000
	142,000		81,000
Válidos	358,000		88,000
Perdidos	,000		35,000
			14,000
			2,000
			47,000
			6,000
		Válidos	358,000
		Perdidos	,000

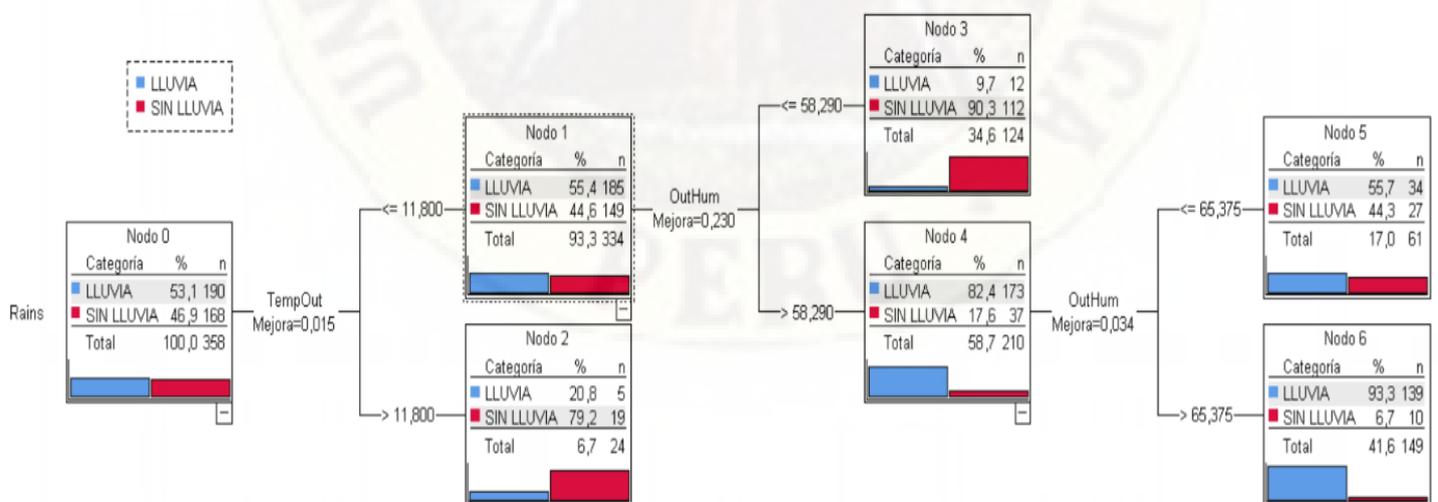
Posteriormente también a partir del modelo obtenido con la base de datos y siguiendo la metodología CRIP-DM se empleó un algoritmo supervisado en la que se usó como variable dependiente o de supervisión a Rains (que indica un día con lluvia o sin lluvia) para predecir que variable meteorológica presenta 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, periodo 2018 – 2019.

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 12**  
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 23**  
Árbol de decisión para la variable dependiente precipitación pluvial-lluvias



**Tabla 13**

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

Observado	Pronosticado			Error	
	LLUVIA	SIN LLUVIA	Porcentaje correcto	Estimación	Desv. Error
LLUVIA	173	17	91,1%	,151	,019
SIN LLUVIA	37	131	78,0%		
Porcentaje global	58,7%	41,3%	84,9%		

Metodo de crecimiento: CRT

Variable dependiente: Rains

Metodo de crecimiento: CRT

Variable dependiente: Rains

## 4.2. Discusión de resultados

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 compresión de los datos (Cruz, Zerquera, Morales, & Rosete, 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, Boza Quintanilla, Herencia Reyes, & Robles Arana, 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, Serna, & Solano, 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).

## 4.3. Proceso de prueba de hipótesis

### 4.3.1. Pruebas de Normalidad

**Paso 1:** Planteamiento de la Hipótesis de Normalidad

**H<sub>0</sub>:** Los datos de las variables meteorológicas en estudio siguen una distribución normal

**H<sub>1</sub>:** Los datos de las variables meteorológicas en estudio no siguen una distribución normal

**Paso 2:** Nivel de confianza

**Nivel de significancia** = 0.95

**$\alpha$  (margen de error)** = 0.05

**Paso 3:** Prueba de normalidad

**Si  $n > 50$**  se aplica Kolmogoron – Smirnov

**Si  $n \leq 50$**  se aplica Shapiro – Wilk

En nuestro caso como tenemos 358 valores por cada variable que son los acumulados y/o promedios diarios en 24 horas, se usa

**Kolmogorov – Smirnov**

**Paso 4:** Estadístico de prueba

**Si  $p\text{-valor} < 0.05$**  se rechaza la H<sub>0</sub>

**Si  $p\text{-valor} \geq 0.05$**  se acepta H<sub>0</sub> y se rechaza la H<sub>1</sub>

**Paso 5:** Criterio de decisión

**Tabla 14**  
*Prueba de Normalidad de las variables meteorológicas en estudio*

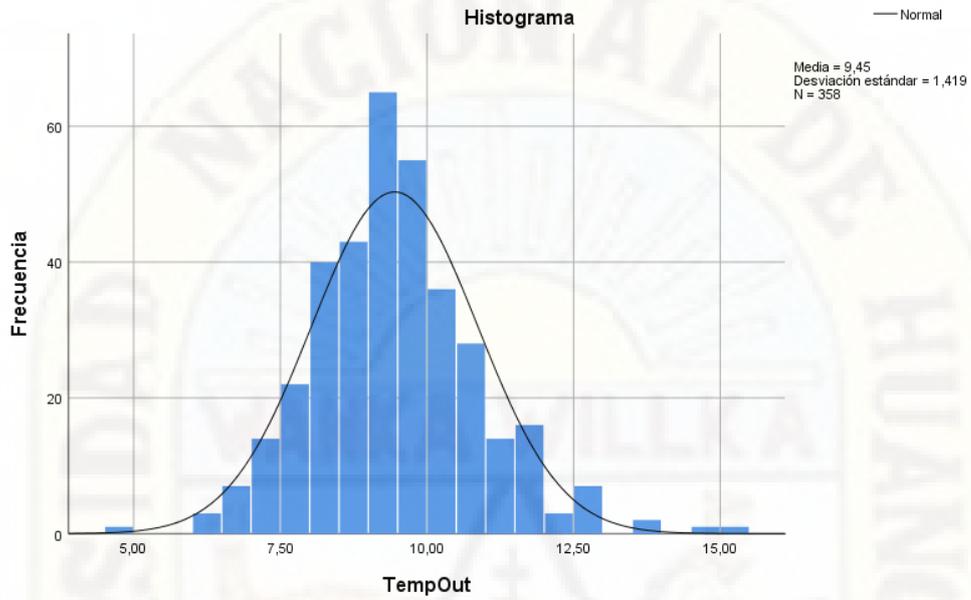
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	gl	p-valor.	Estadístico	gl	p-valor
TempOut	,063	358	<b>,002</b>	,981	358	,000
OutHum	,078	358	<b>,000</b>	,958	358	,000
BarmmHg	,332	358	<b>,000</b>	,733	358	,000
SolarRad	,049	358	<b>,038</b>	,988	358	,006
UVIndexmáx	,077	358	<b>,000</b>	,956	358	,000
WindSpeedmáx	,153	358	<b>,000</b>	,950	358	,000
Rain	,286	358	<b>,000</b>	,624	358	,000

Corrección de significación de Lilliefors

Teniendo en cuenta que nuestros datos son mayores a 50 (358) se utilizó los resultados de la prueba de Kolmogórov-Smirnov en la que se observa en la Tabla 14 que el p-valor de las variables meteorológicas en estudio en todos los casos son menores que **0.05** entonces se rechaza la **H<sub>0</sub>** y se acepta la **H<sub>1</sub>**, esto quiere decir que en todos los casos las siete variables meteorológicas no siguen una distribución normal, lo que se verifica con los gráficos de normalidad que se presenten en las figuras siguientes:

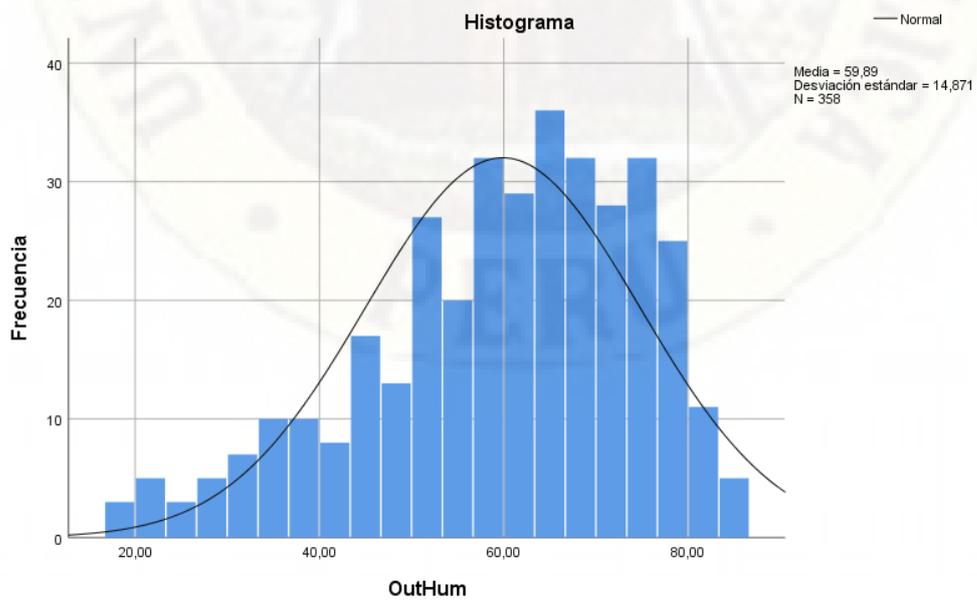
**Figura 24**

*Curva de normalidad de la variable temperatura ambiental promedio diario*

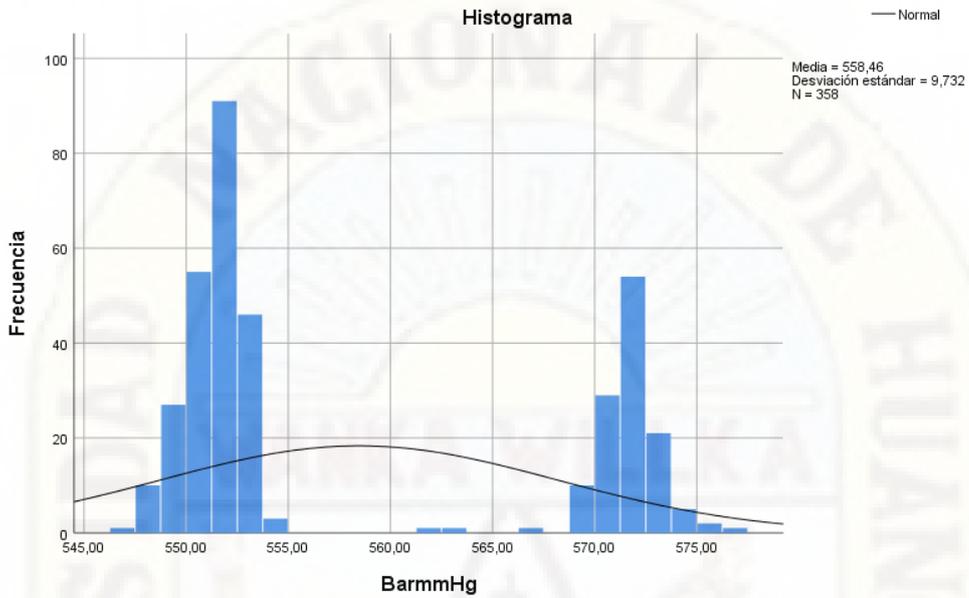


**Figura 25**

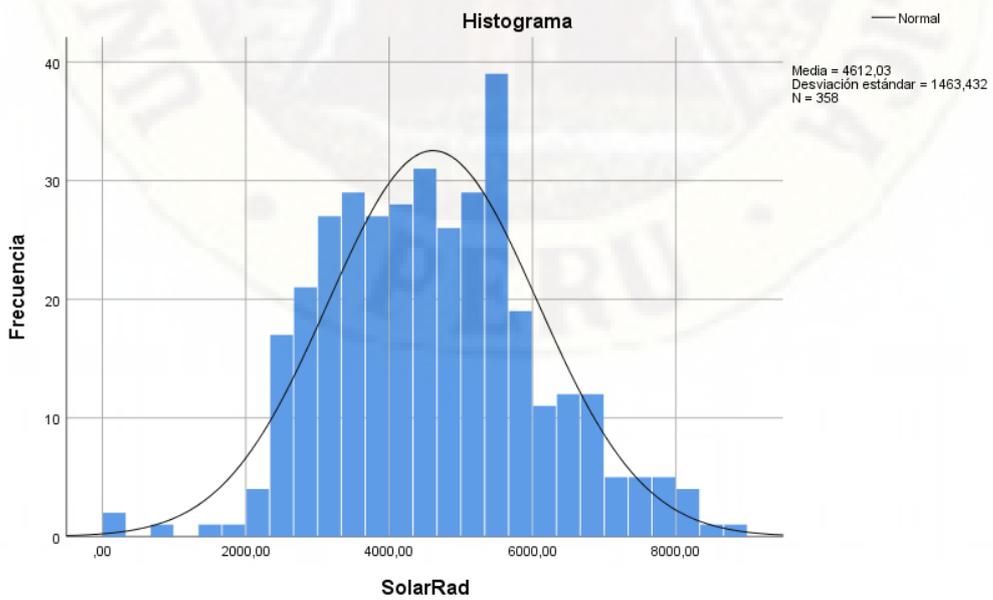
*Curva de normalidad de la variable humedad relativa promedio diario*



**Figura 26**  
*Curva de normalidad de la variable presión atmosférica promedio diario*

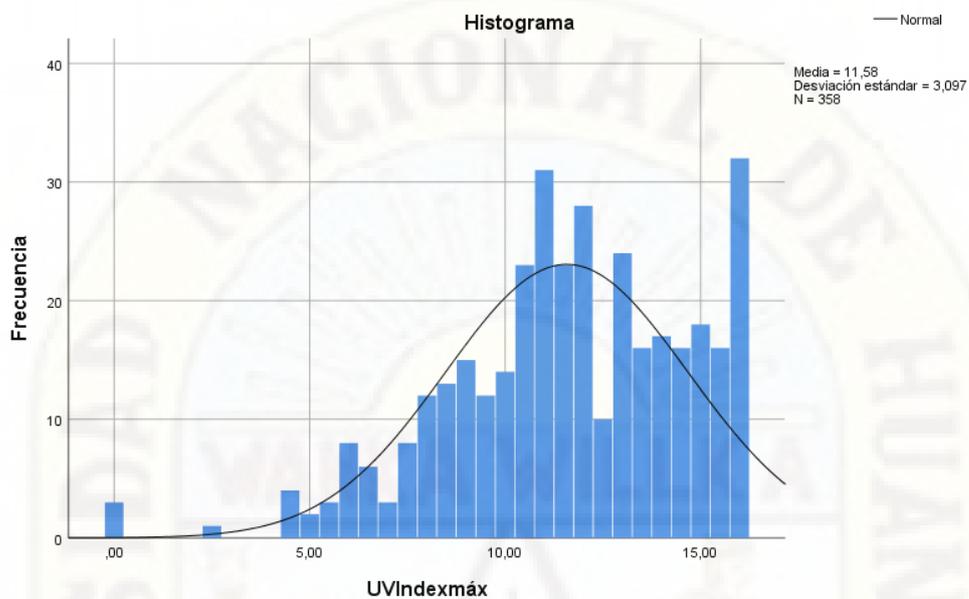


**Figura 27**  
*Curva de normalidad de la variable Radiación solar acumulada diario*



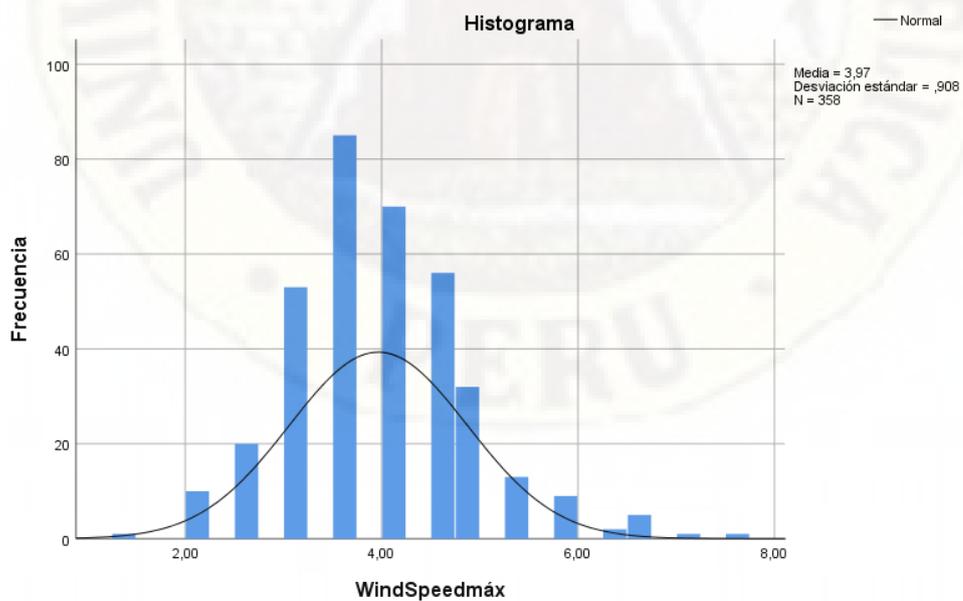
**Figura 28**

*Curva de normalidad de la variable índice de radiación ultravioleta máxima diaria*



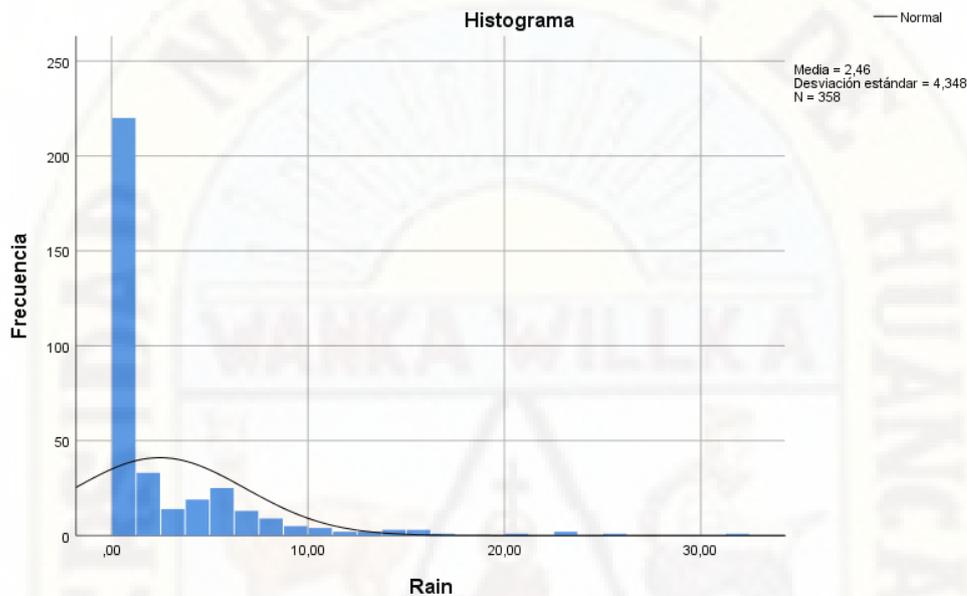
**Figura 29**

*Curva de normalidad de la variable velocidad del viento máxima diaria*



**Figura 30**

*Curva de normalidad de la variable precipitación pluvial acumulada diaria*



#### 4.3.2. Pruebas de Hipótesis

##### a) Prueba de la Hipótesis específica 1

###### Paso 1: Planteamiento de las Hipótesis estadísticas

**H<sub>0</sub>:** Los clústeres formados mediante técnicas de Data Mining, a partir de valores meteorológicos diarios en la ciudad de Huancavelica, 2018–2019, no presentan diferencias significativas

**H<sub>1</sub>:** Los clústeres formados mediante técnicas de Data Mining, a partir de valores meteorológicos diarios en la ciudad de Huancavelica, 2018–2019, presentan diferencias significativas

**Paso 2:** Nivel de confianza

Nivel de significancia = 0.95

$\alpha$  (margen de error) = 0.05

**Paso 3:** Prueba de significancia

**Paramétrica** -> Coeficiente de correlación de Pearson

**No paramétrica** -> Algoritmos de segmentación K-means

En nuestro caso como las pruebas de Normalidad indicaron que las variables de estudio no presentan Normalidad, entonces se usó una técnica no paramétrica, particularmente para este proceso se usó una técnica de Data Mining denominado segmentación, usando el algoritmo no supervisado de K-means.

**Paso 4:** Estadístico de prueba

Si **p-valor < 0.05** se rechaza la **H<sub>0</sub>**

Si **p-valor ≥ 0.05** se acepta **H<sub>0</sub>** y se rechaza la **H<sub>1</sub>**

**Paso 5:** Criterio de decisión

**Tabla 15**

Tabla de ANOVA respecto a cuatro clústeres formados mediante la técnica de Data Mining denominada segmentación con el algoritmo no supervisado k-means

	Clúster		Error		F	p-valor
	Media cuadrática	gl	Media cuadrática	gl		
TempOut	57,188	3	1,545	354	37,012	,000
Rain	68,566	3	18,483	354	3,710	,012
OutHum	8257,296	3	153,044	354	53,954	,000
BarmmHg	1711,735	3	81,011	354	21,130	,000
SolarRad	223588395,689	3	264966,954	354	843,835	,000
UVIndexmáx	543,218	3	5,069	354	107,172	,000
WindSpeedmáx	10,173	3	,745	354	13,652	,000

Las pruebas F sólo se deben utilizar con fines descriptivos porque los clústeres se han elegido para maximizar las diferencias entre los casos de distintos clústeres.

**Tabla 16**

Tabla de ANOVA respecto a diez clústeres formados mediante la técnica de Data Mining denominada segmentación con el algoritmo no supervisado k-means

ANOVA

	Clúster		Error		F	p-valor
	Media cuadrática	gl	Media cuadrática	gl		
TempOut	26,056	9	1,391	348	18,734	,000
Rain	44,277	9	18,247	348	2,426	,011
OutHum	3504,735	9	136,227	348	25,727	,000
BarmmHg	660,167	9	80,091	348	8,243	,000
SolarRad	82567621,303	9	61652,003	348	1339,253	,000
UVIndexmáx	217,940	9	4,203	348	51,858	,000
WindSpeedmáx	4,540	9	,728	348	6,233	,000

Las pruebas F sólo se deben utilizar con fines descriptivos porque los clústeres se han elegido para maximizar las diferencias entre los casos de distintos clústeres.

Como el p-valor tanto en el análisis de varianzas para 4 y seis grupos o clúster formados son menores a **0.05** para todas las variables meteorológicas estudiadas, entonces se rechaza **H<sub>0</sub>**, por tanto, se acepta **H<sub>1</sub>**, esto quiere decir que “Los clústeres formados mediante Técnicas de Data Mining, a partir de valores meteorológicos diarios en la ciudad de Huancavelica, 2018–2019, presentan diferencias significativas”

## b) Prueba de la Hipótesis específica 2

### Paso 1: Planteamiento de las Hipótesis estadísticas

**H<sub>0</sub>:** Las variables meteorológicas procesados mediante técnicas de Data Mining en los diferentes meses del año en la ciudad de Huancavelica, 2018 – 2019, no presentan diferencias significativas.

**H<sub>1</sub>:** Las variables meteorológicas procesados mediante técnicas de Data Mining en los diferentes meses del año en la ciudad de Huancavelica, 2018 – 2019, presentan diferencias significativas.

**Paso 2:** Nivel de confianza

Nivel de significancia = 0.95

$\alpha$  (margen de error) = 0.05

**Paso 3:** Prueba de significancia

**Paramétrica** -> Coeficiente de correlación de Pearson

**No paramétrica**-> Prueba de Kruskal-Walis

En nuestro caso como las pruebas de Normalidad indicaron que las variables de estudio no presentan Normalidad, entonces se usó una técnica no paramétrica, particularmente para este proceso se usó una técnica de Data Mining denominado segmentación, usando el algoritmo no supervisado de K-means.

**Paso 4:** Estadístico de prueba

Si **p-valor** < **0.05** se rechaza la **H<sub>0</sub>**

Si **p-valor** ≥ **0.05** se acepta **H<sub>0</sub>** y se rechaza la **H<sub>1</sub>**

**Paso 5:** Criterio de decisión

**Tabla 17**

Tabla de la prueba de hipótesis respecto las variables meteorológicas en meses, ciudad de Hvca., 2018 – 2019

Variable	Ago	Set	Oct	Nov	Dic	Ene	Feb	Mar	Abr	May	Jun	Jul	Prueba F	gl	Importancia	p-valor
UVIndexmÁix	11.16	12.54	11.36	14.02	12.80	12.12	12.48	11.23	11.12	10.45	9.86	10.28	4.83	11,346	1.00	0.00
	2.26	3.03	4.14	2.84	3.97	3.11	2.74	2.57	3.56	2.28	1.80	1.80			Importante	
	0.41	0.55	0.76	0.56	0.74	0.56	0.52	0.46	0.65	0.41	0.33	0.32				
	31.00	30.00	30.00	26.00	29.00	31.00	28.00	31.00	30.00	31.00	30.00	31.00				
SolarRad	4156.81	5340.37	4248.20	5578.92	5550.59	4673.71	4234.71	3861.74	4099.30	4249.26	4710.37	4818.71	5.36	11,346	1.00	0.00
	849.14	1529.07	1745.71	1937.22	1468.20	1676.13	1212.66	980.93	1420.60	1226.36	1078.15	1017.66			Importante	
	152.51	279.17	318.72	379.92	272.64	301.04	229.17	176.18	259.37	220.26	196.84	182.78				
	31.00	30.00	30.00	26.00	29.00	31.00	28.00	31.00	30.00	31.00	30.00	31.00				
WindSpeedmÁix	3.74	4.01	4.01	4.29	5.00	3.92	4.28	3.63	3.87	3.50	3.71	3.78	6.75	11,346	1.00	0.00
	0.85	0.89	1.15	0.79	0.88	1.04	0.86	0.66	0.92	0.67	0.62	0.51			Importante	
	0.15	0.16	0.21	0.15	0.16	0.19	0.16	0.12	0.17	0.12	0.11	0.09				
	31.00	30.00	30.00	26.00	29.00	31.00	28.00	31.00	30.00	31.00	30.00	31.00				
Rain	1.28	0.68	3.59	1.61	1.37	5.80	5.96	6.34	1.55	1.11	0.01	0.17	11.54	11,346	1.00	0.00
	2.32	1.79	4.34	3.05	2.21	6.38	6.42	6.00	2.40	2.40	0.05	0.65			Importante	
	0.42	0.33	0.79	0.60	0.41	1.15	1.21	1.08	0.44	0.43	0.01	0.12				
	31.00	30.00	30.00	26.00	29.00	31.00	28.00	31.00	30.00	31.00	30.00	31.00				
TempOut	7.81	9.37	9.57	11.35	10.60	10.06	9.44	9.68	9.80	9.34	8.43	8.29	23.30	11,346	1.00	0.00
	0.82	1.71	0.77	1.20	1.45	1.51	1.03	0.68	0.79	0.81	1.02	0.76			Importante	
	0.15	0.31	0.14	0.24	0.27	0.27	0.19	0.12	0.14	0.15	0.19	0.14				
	31.00	30.00	30.00	26.00	29.00	31.00	28.00	31.00	30.00	31.00	30.00	31.00				
OutHum	63.80	52.85	66.86	57.48	55.85	69.45	74.84	73.52	64.04	55.33	43.15	41.87	29.99	11,346	1.00	0.00
	7.60	11.56	9.91	13.99	15.11	11.56	6.03	4.92	9.11	11.70	11.58	12.31			Importante	
	1.37	2.11	1.81	2.74	2.81	2.08	1.14	0.88	1.66	2.10	2.11	2.21				
	31.00	30.00	30.00	26.00	29.00	31.00	28.00	31.00	30.00	31.00	30.00	31.00				
BarrmmHg	571.78	554.22	551.26	549.51	551.08	551.29	551.86	551.88	552.20	569.37	572.38	571.97	481.93	11,346	1.00	0.00
	1.07	5.93	0.77	1.12	1.46	1.64	1.03	0.70	0.76	4.66	1.70	1.10			Importante	
	0.19	1.08	0.14	0.22	0.27	0.29	0.20	0.13	0.14	0.84	0.31	0.20				
	31.00	30.00	30.00	26.00	29.00	31.00	28.00	31.00	30.00	31.00	30.00	31.00				

Fuente: Propia, elaborado en IBM - SPSS-Modeler Subscription

Debido a que las variables no presentan una distribución normal se aplicó una prueba no paramétrica denominada Kruskal-Wallis para realizar la prueba de hipótesis y como el p-valor para todas las variables analizadas son menores a **0.05**, entonces se rechaza **H<sub>0</sub>**, por tanto, se acepta **H<sub>1</sub>**, esto quiere decir que “Las variables meteorológicas procesados mediante técnicas de Data Mining en los diferentes meses del año en la ciudad de Huancavelica, 2018 – 2019, presentan diferencias significativas.”



## 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 de 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.

## **Recomendaciones**

Los estudios referidos a las variables meteorológicas son cada vez más frecuentes debido a los impactos del cambio climático, en la presente investigación se realizó un análisis global de las diferentes variables meteorológicas que pueda registrar una estación meteorológica automática, así como también de las distintas técnicas y algoritmos propios de Data Mining, se sugiere en adelante realizar investigaciones del uso de estos algoritmos usando una o dos variables y usando una determinada técnica o algoritmo computacional, a fin de precisar las conclusiones y determinar los procedimientos de uso y manejo de estas herramientas para un adecuado monitoreo del cambio climático y si fuera posible incrementar los porcentajes de probabilidad de éxito en las predicciones.

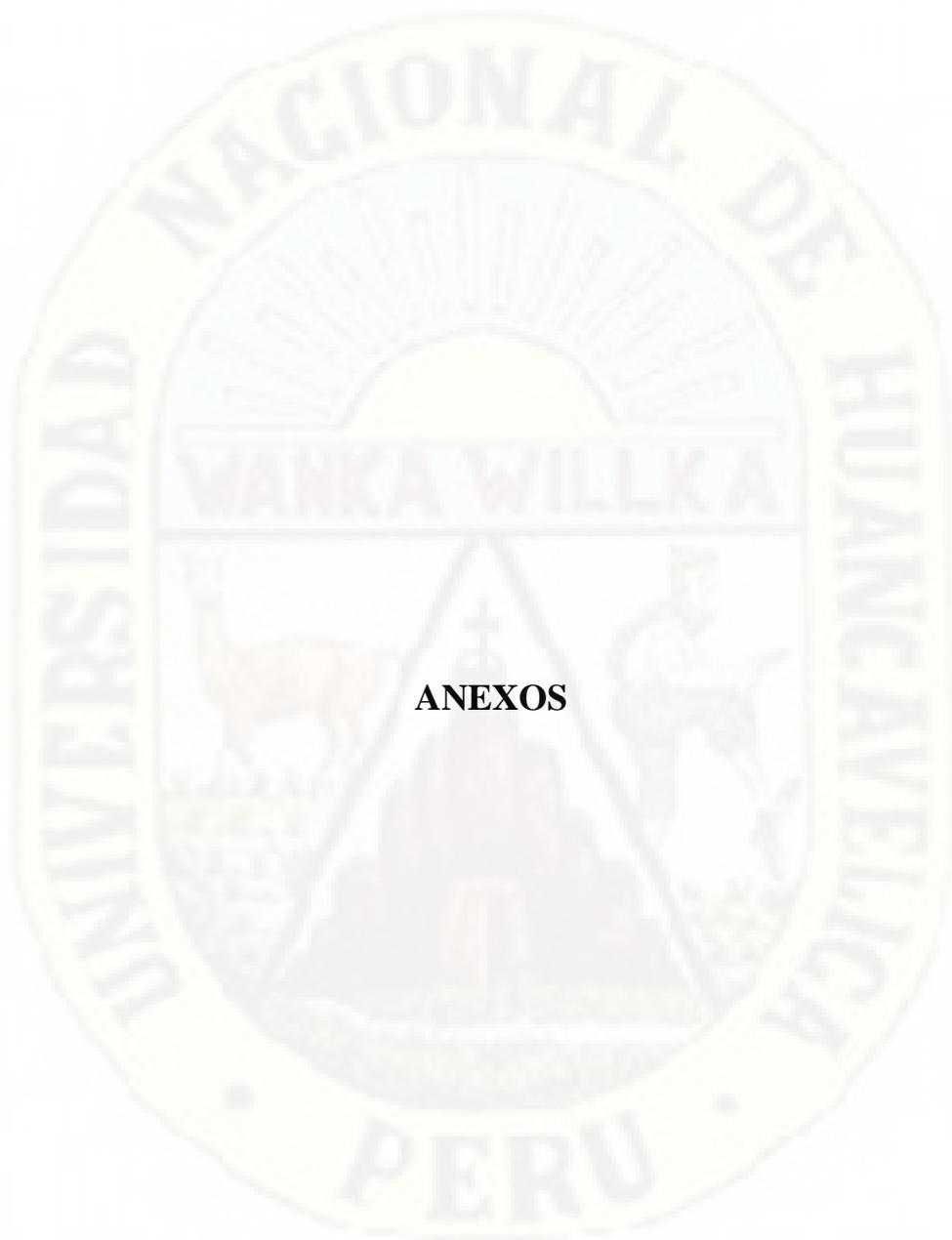
Los resultados obtenidos en la presente investigación, son una base de futuras investigaciones, pero se debe tener en cuenta que estos cambios de comportamiento y patrones para una posible predicción pueden cambiar significativamente dependiendo del lugar y los puntos de obtención de los datos, así como también de los instrumentos utilizados para obtener los datos, y las diferentes y variadas metodologías usadas para las etapas del procesamiento de los datos teniendo en cuenta un buen entendimiento del problema, entendimiento de los datos, preparación y limpieza de los datos, modelos adecuados a usarse y una adecuada evaluación, interpretación de los resultados obtenidos, tener en cuenta para futuras investigaciones en este campo de la meteorología y monitoreo ambiental,

## Referencias Bibliográficas

- 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/IJARCC.2016.54266
- Mejía, 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. Valparaíso: Pontificia Universidad Católica de Valparaíso.

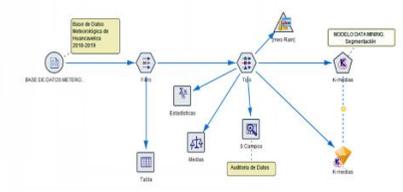


**ANEXOS**

# MATRIZ DE CONSISTENCIA

Título:

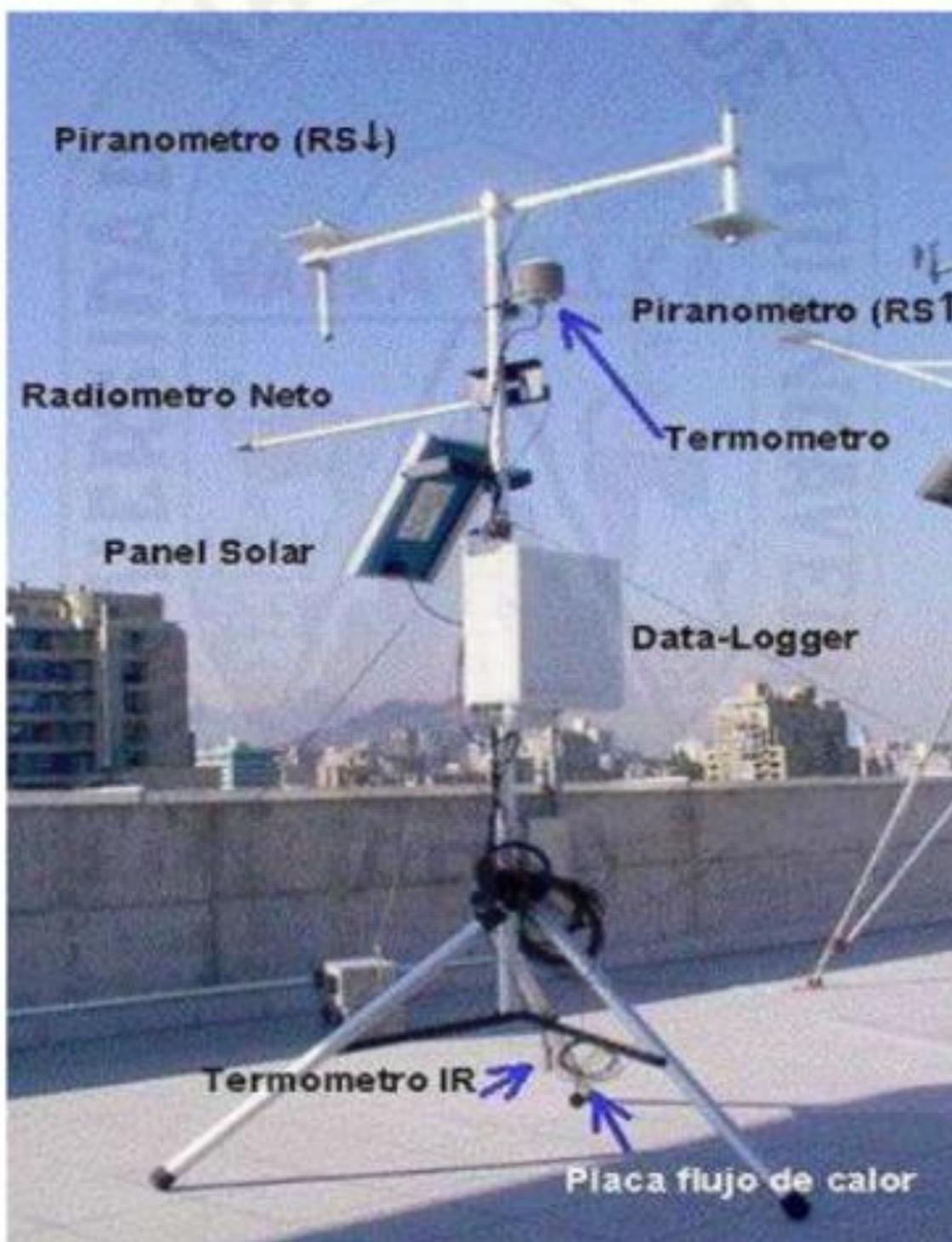
**PATRONES DE COMPORTAMIENTO DE DATOS METEOROLÓGICOS MEDIANTE TÉCNICAS DE DATA MINING EN HUANCVELICA 2018 – 2019**

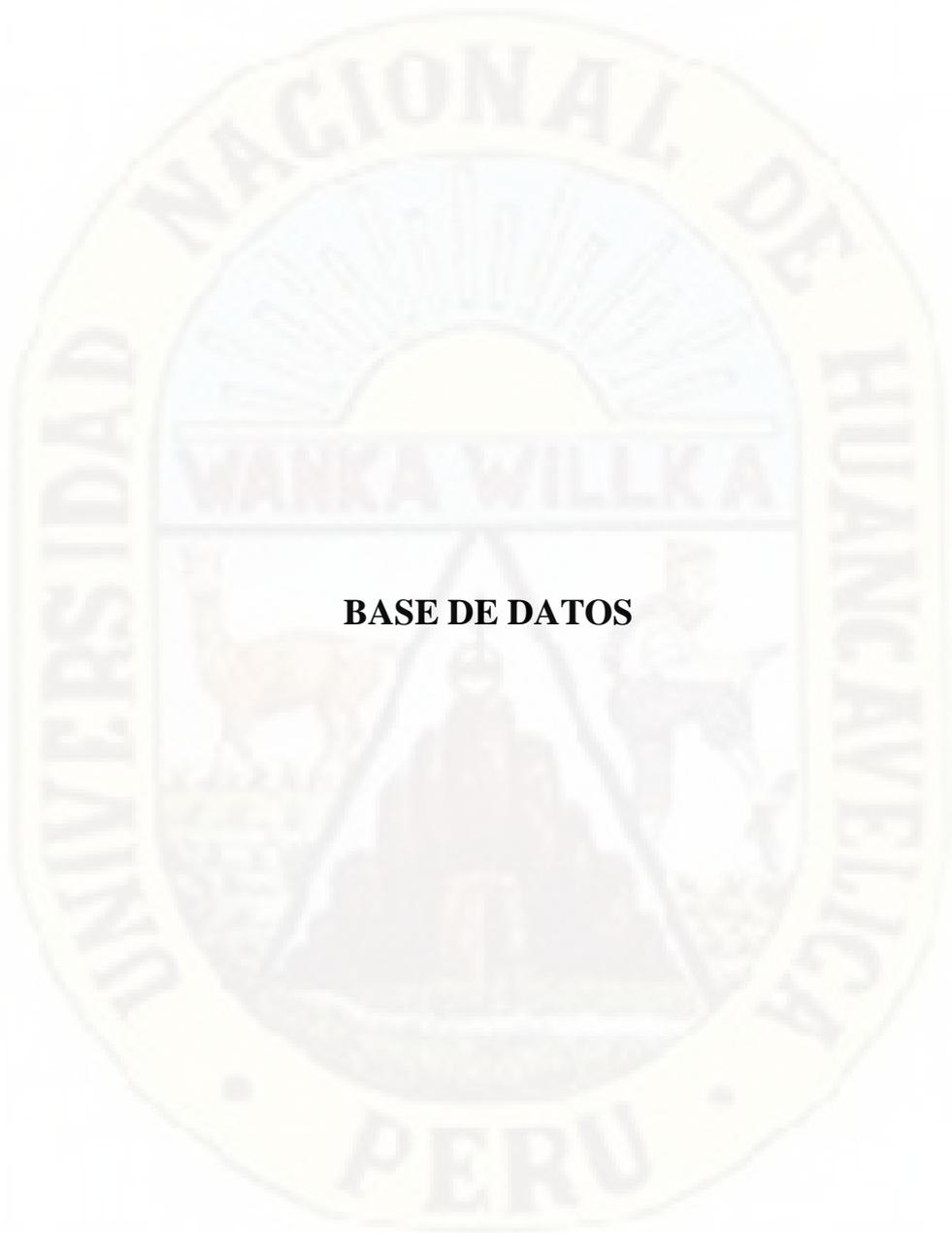
FORMULACION DEL PROBLEMA	OBJETIVOS	HIPOTESIS	VARIABLES	DIMENSIONES	MUESTRA Y MUESTREO	METODOLOGIA
<p><b>Problema General</b></p> <p>¿Cuáles son los patrones de comportamiento obtenidos mediante técnicas de Data Mining, de datos asociados a las variables meteorológicas en la ciudad de Huancavelica, 2018 - 2019?</p> <p><b>Problemas Específicos</b></p> <p>a) ¿Existen diferencias significativas entre clúster formados mediante técnicas de Data Mining, a partir de valores meteorológicos diarios en la ciudad de Huancavelica, 2018 - 2019?</p> <p>b) ¿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?</p>	<p><b>Objetivo General</b></p> <p>Determinar patrones de comportamiento de datos obtenidos mediante técnicas de Data Mining, asociados a las variables meteorológicas en la ciudad de Huancavelica, 2018 – 2019.</p> <p><b>Objetivos Específicos</b></p> <p>a) 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.</p> <p>b) 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.</p>	<p><b>Hipótesis General</b></p> <p>Los patrones de comportamiento a partir de datos de variables meteorológicas pueden ser obtenidos mediante técnicas de Data Mining en la ciudad de Huancavelica, 2018 – 2019.</p> <p><b>Hipótesis Específicas</b></p> <p>a) Los clústeres formados mediante técnicas de Data Mining, a partir de valores meteorológicos diarios en la ciudad de Huancavelica, 2018–2019, presentan diferencias significativas</p> <p>b) Las variables meteorológicas procesados mediante técnicas de Data Mining y los diferentes meses del año en la ciudad de Huancavelica, 2018 – 2019, presentan diferencia significativa.</p>	<p><b>Variable 1</b></p> <p>Variables meteorológicas en la ciudad de Huancavelica 2018–2019.</p> <ul style="list-style-type: none"> <li>▪ Temperatura Ambiental</li> <li>▪ Precipitación Pluvial</li> <li>▪ Humedad relativa</li> <li>▪ Presión atmosférica</li> <li>▪ Radiación Solar</li> <li>▪ Radiación Ultravioleta</li> <li>▪ Velocidad del viento</li> </ul>		<p><b>MUESTRA</b></p> <p>Momentos horarios en las que se evaluó las variables meteorológicas que estuvo conformada por el periodo de estudio muestral de un año establecido del primero de agosto del 2018 al 31 de julio del 2019</p> <p><b>MUESTREO</b></p> <p>La técnica de muestreo del presente trabajo fue el muestreo aleatorio sistémico, ya que la estación meteorológica estuvo programada para registrar datos meteorológicos de a ser procesados sistemáticamente cada 60 min. durante 12 meses y 358 días.</p>	<p><b>Tipo, Nivel y Método</b></p> <p>El presente trabajo es de tipo Descriptivo, nivel básico y se usó para su desarrollo el método científico</p> <p><b>Diseño</b></p> <p>En la presente investigación se usó el siguiente Diseño de investigación longitudinal;</p>  <p><b>Técnicas de análisis de resultados</b></p> <ul style="list-style-type: none"> <li>• Estadística descriptiva</li> <li>• Estadística no paramétrica</li> <li>• Data Mining</li> </ul>

## INSTRUMENTO DE RECOLECCIÓN DE DATOS

Figura 31

*Instrumento de medición de variables meteorológicas (EMA) Estación Meteorológica Automatizada*





**BASE DE DATOS**

UNIVERSIDAD NACIONAL DE HUANCAMELICA - FACULTAD DE CIENCIAS DE INGENIERIA - BASE DE DATOS  
 UNIDAD DE POSGRADO - DOCTORADO EN CIENCIAS AMBIENTALES - M.Sc. Jorge Luis HUERE PEÑA

ORD	año	mes	día	Temp Out	Rain	Out Hum	Bar mmHg	Solar Rad	UV Index máx	Speed máx	Rains
1	2018	Ago	1	7.73	4.57	65.67	570.03	3674	11	4	LLUVIA
2	2018	Ago	2	7.96	4.83	68.75	571.18	5067	11.6	3.6	LLUVIA
3	2018	Ago	3	7.25	0.51	74.42	570.18	3217	7.8	4.5	LLUVIA
4	2018	Ago	4	8.1	0	59.79	570.83	5165	12.2	4	SIN LLUVIA
5	2018	Ago	5	7.17	0	51.54	570.93	5517	13.2	4	SIN LLUVIA
6	2018	Ago	6	7.22	0.5	65.54	571.09	3778	11.3	4	LLUVIA
7	2018	Ago	7	6.56	10.15	75.21	572.3	3066	9	4.5	LLUVIA
8	2018	Ago	8	6.06	0.76	74.3	574.11	3269	9.7	3.6	LLUVIA
9	2018	Ago	9	7.8	0.5	70.42	574.07	3392	10.3	3.1	LLUVIA
10	2018	Ago	10	7.93	1.27	67.5	571.93	3861	12.6	3.1	LLUVIA
11	2018	Ago	11	7.78	0	63.08	573.02	4141	9.3	2.2	SIN LLUVIA
12	2018	Ago	12	8.68	0	57.54	572.39	4881	7.7	3.1	SIN LLUVIA
13	2018	Ago	13	7.09	0	58	571.97	3740	10.3	3.6	SIN LLUVIA
14	2018	Ago	14	7.06	0	51.21	572.41	4497	10.7	4	SIN LLUVIA
15	2018	Ago	15	6.74	0.25	61	572.54	3854	10.7	4	LLUVIA
16	2018	Ago	16	7.8	0.25	64.42	573.12	2899	6.5	4.5	LLUVIA
17	2018	Ago	17	8.32	0.25	66.54	572.51	2690	7.7	4.5	LLUVIA
18	2018	Ago	18	7.64	2.53	69.75	571.81	4445	12.9	3.6	LLUVIA
19	2018	Ago	19	7.63	1.52	69.92	571.69	4528	13.2	4.9	LLUVIA
20	2018	Ago	20	8.46	1.52	69.67	571.47	3817	11.1	3.6	LLUVIA
21	2018	Ago	21	8.53	0.25	69.29	570.83	3248	10.5	3.6	LLUVIA
22	2018	Ago	22	8.14	1.77	70.46	571.45	4327	14.2	3.1	LLUVIA
23	2018	Ago	23	7.8	0	59.08	572.02	4808	14.2	2.7	SIN LLUVIA
24	2018	Ago	24	8.17	0	53.5	571.25	5208	13.9	3.6	SIN LLUVIA
25	2018	Ago	25	7.74	0	57.92	572.06	3763	8.4	4	SIN LLUVIA
26	2018	Ago	26	8.02	6.85	67.04	571.59	4799	11.9	6.7	LLUVIA
27	2018	Ago	27	6.9	1.26	74.67	572.29	2913	9	3.6	LLUVIA
28	2018	Ago	28	7.79	0.25	63.96	572.56	4642	12.2	3.6	LLUVIA
29	2018	Ago	29	8.53	0	56.21	571.56	4911	14.8	3.6	SIN LLUVIA
30	2018	Ago	30	9.39	0	50	570.45	4742	14.1	2.2	SIN LLUVIA
31	2018	Ago	31	10.25	0	51.25	569.41	6002	13.9	2.7	SIN LLUVIA
32	2018	Set	1	9.51	0	49.25	569.15	5241	12.8	3.6	SIN LLUVIA
33	2018	Set	2	9.23	0	39.71	570.81	6562	13.9	4.9	SIN LLUVIA
34	2018	Set	3	8.27	0	56.58	570.89	4284	10.1	3.1	SIN LLUVIA
35	2018	Set	4	8.45	0	48.92	563.37	6663	14.4	3.1	SIN LLUVIA
36	2018	Set	5	8.92	0	42.21	553.31	6303	14.2	3.6	SIN LLUVIA
37	2018	Set	6	9.17	0	32.88	553.33	6740	14.6	3.6	SIN LLUVIA
38	2018	Set	7	9.73	0	31.29	553.36	6798	14	4.9	SIN LLUVIA
39	2018	Set	8	9.66	0	37.96	552.84	6137	13.6	4.5	SIN LLUVIA
40	2018	Set	9	9.92	0	44.71	552.31	5574	12.1	3.6	SIN LLUVIA
41	2018	Set	10	9.38	0	53.71	552.48	4451	11.2	2.7	SIN LLUVIA
42	2018	Set	11	9.82	0	50.46	552.49	4431	11	2.7	SIN LLUVIA
43	2018	Set	12	9.82	0	50.13	551.98	5346	13.2	4.9	SIN LLUVIA
44	2018	Set	13	10.69	0	45.13	550.67	6090	13.7	4.9	SIN LLUVIA

UNIVERSIDAD NACIONAL DE HUANCAMELICA - FACULTAD DE CIENCIAS DE INGENIERIA - BASE DE DATOS  
 UNIDAD DE POSGRADO - DOCTORADO EN CIENCIAS AMBIENTALES - M.Sc. Jorge Luis HUERE PEÑA

ORD	año	mes	día	Temp Out	Rain	Out Hum	Bar mmHg	Solar Rad	UV Index máx	Speed máx	Rains
45	2018	Set	14	10.06	0	52.67	550.21	4433	9.4	3.6	SIN LLUVIA
46	2018	Set	15	8.48	2.79	62.04	550.76	3673	8.9	2.2	LLUVIA
47	2018	Set	16	9	2.29	60.83	552.07	5140	12.3	4.9	LLUVIA
48	2018	Set	17	9.42	0	53.19	552.24	5839	15.7	4	SIN LLUVIA
49	2018	Set	18	6.3	0	55	553.43	3136	13.3	4	SIN LLUVIA
50	2018	Set	19	14.81	0	33.94	550.42	8716	15.8	4.5	SIN LLUVIA
51	2018	Set	20	10.14	0	52.92	550.75	6702	15.2	3.6	SIN LLUVIA
52	2018	Set	21	9.16	0	58.56	552.26	5785	15.6	4.5	SIN LLUVIA
53	2018	Set	22	4.73	8.38	85.57	552.6	10	0	2.7	LLUVIA
54	2018	Set	23	8.42	0	64.71	553.14	5877	14.7	5.8	SIN LLUVIA
55	2018	Set	24	7.9	4.57	68.29	552.31	4280	9.4	4	LLUVIA
56	2018	Set	25	7.87	0.25	63.71	553.27	5607	13.3	4.9	LLUVIA
57	2018	Set	26	8.66	0	61.92	553.48	4133	12.1	5.4	SIN LLUVIA
58	2018	Set	27	10.18	1.02	62.96	550.98	4598	11.4	4.5	LLUVIA
59	2018	Set	28	10.48	1.02	57.67	551.03	6105	14.6	4.5	LLUVIA
60	2018	Set	29	11.05	0	55.25	550.61	5286	12.9	3.1	SIN LLUVIA
61	2018	Set	30	11.92	0	53.42	550.15	6271	12.9	4	SIN LLUVIA
62	2018	Oct	1	9.92	0	61.47	550.27	2542	5.5	4	SIN LLUVIA
63	2018	Oct	2	8.97	0	65.17	551.4	2	0	1.3	SIN LLUVIA
64	2018	Oct	3	10.13	0	60.67	551.41	2866	7.9	4.5	SIN LLUVIA
65	2018	Oct	5	9.41	0	67.08	551.52	2171	4.7	4.5	SIN LLUVIA
66	2018	Oct	6	9.85	1.53	67.04	551.46	4317	13.1	2.7	LLUVIA
67	2018	Oct	7	9.01	2.54	71.58	551.7	4271	10.2	4.5	LLUVIA
68	2018	Oct	8	9.9	0	47.08	552.8	6527	14.7	4.5	SIN LLUVIA
69	2018	Oct	9	10.52	0	45.75	550.72	6429	15	3.1	SIN LLUVIA
70	2018	Oct	10	10.68	0	53.59	549.52	2838	9.3	4.5	SIN LLUVIA
71	2018	Oct	11	9.48	0.51	57.5	550.97	4048	10.9	4	LLUVIA
72	2018	Oct	12	10.37	0	51	551.88	6975	15.4	4	SIN LLUVIA
73	2018	Oct	13	9.21	5.07	73.75	551.09	2729	8.4	3.1	LLUVIA
74	2018	Oct	14	8.58	5.84	79.5	552.42	2651	6.1	3.1	LLUVIA
75	2018	Oct	15	9.42	0.76	70.79	552.31	2836	5.9	3.6	LLUVIA
76	2018	Oct	16	9.53	1.02	69.21	551.13	4141	14	4.5	LLUVIA
77	2018	Oct	17	11.11	5.85	53.29	550.93	8335	16	3.6	LLUVIA
78	2018	Oct	18	10.25	8.64	65.79	549.99	6453	15	4.9	LLUVIA
79	2018	Oct	19	9.13	7.87	64.63	551.35	4965	13.8	7.6	LLUVIA
80	2018	Oct	20	8.42	6.6	71.63	551.81	4841	14.3	5.8	LLUVIA
81	2018	Oct	21	7.99	4.82	81.21	551.33	3383	8.3	4.5	LLUVIA
82	2018	Oct	22	8.18	7.09	79.08	552.16	3352	7.6	3.6	LLUVIA
83	2018	Oct	23	9.86	5.33	73.04	550.82	4285	13.5	4.9	LLUVIA
84	2018	Oct	24	10.08	1.26	77.67	550.07	4236	15.4	3.1	LLUVIA
85	2018	Oct	25	9.17	4.58	79.25	550.36	3743	14.9	2.7	LLUVIA
86	2018	Oct	26	8.59	0.51	77.78	551.43	4030	13.4	3.6	LLUVIA
87	2018	Oct	27	9.37	6.09	74.92	551.43	3759	11	4	LLUVIA
88	2018	Oct	28	10.81	0	65.21	551.16	6733	16	4.5	SIN LLUVIA

UNIVERSIDAD NACIONAL DE HUANCAMELICA - FACULTAD DE CIENCIAS DE INGENIERIA - BASE DE DATOS  
 UNIDAD DE POSGRADO - DOCTORADO EN CIENCIAS AMBIENTALES- M.Sc. Jorge Luis HUERE PEÑA

ORD	año	mes	día	Temp Out	Rain	Out Hum	Bar mmHg	Solar Rad	UV Index máx	Speed máx	Rains
89	2018	Oct	29	10	20.58	59.42	550.36	4413	12.8	2.7	LLUVIA
90	2018	Oct	30	9.82	5.33	67.5	551.86	6133	15.9	5.4	LLUVIA
91	2018	Oct	31	9.29	5.85	74.33	552.16	3442	11.8	3.6	LLUVIA
92	2018	Nov	1	8.73	1.01	75.88	552.46	3657	10	4	LLUVIA
93	2018	Nov	2	9.5	13.7	74.5	551.36	5591	15.8	4.5	LLUVIA
94	2018	Nov	3	9.93	4.32	69.63	551.1	7462	16	5.4	LLUVIA
95	2018	Nov	4	9.93	0.25	67.04	549.92	6892	16	4.5	LLUVIA
96	2018	Nov	5	11.25	0	56.83	549.75	6924	15.7	3.6	SIN LLUVIA
97	2018	Nov	6	12.07	0	48.54	548.8	6662	13	4	SIN LLUVIA
98	2018	Nov	7	11.97	0	61.75	548.26	4151	9.2	3.6	SIN LLUVIA
99	2018	Nov	8	11.42	0.51	64.67	548.74	4010	11.9	4	LLUVIA
100	2018	Nov	9	11.55	7.11	58.63	549.91	6683	14.7	3.6	LLUVIA
101	2018	Nov	10	11.89	1.77	61.04	548.79	7133	15.2	3.6	LLUVIA
102	2018	Nov	11	10.6	2.54	66.21	548.79	4110	9.8	3.6	LLUVIA
103	2018	Nov	12	11.64	0	56.42	549.44	5659	15.3	4	SIN LLUVIA
104	2018	Nov	13	11.97	0	32.3	549.75	6982	16	3.7	SIN LLUVIA
105	2018	Nov	14	12.49	0	25.85	549.05	709	16	5.3	SIN LLUVIA
106	2018	Nov	15	11.78	4.4	60.36	547.79	4248	12.3	4.7	LLUVIA
107	2018	Nov	16	11.12	0	62.42	549.93	5717	15.8	5.4	SIN LLUVIA
108	2018	Nov	17	12.98	0	37.64	549.79	8247	16	4.5	SIN LLUVIA
109	2018	Nov	18	10.82	0.4	52.7	549.34	5448	13.5	5.8	LLUVIA
110	2018	Nov	19	10.51	3	65.04	550.42	5871	15.8	5.8	LLUVIA
111	2018	Nov	20	12.1	0	63.29	550.15	6626	16	4.9	SIN LLUVIA
112	2018	Nov	21	11.57	0	65.14	549.52	3683	13.5	3.6	SIN LLUVIA
113	2018	Nov	22	12.82	0	55.71	549.3	7892	15.9	4.5	SIN LLUVIA
114	2018	Nov	23	13.8	0	32.46	548.06	8042	16	3.6	SIN LLUVIA
115	2018	Nov	24	12.57	0	36.21	547.75	7402	16	4	SIN LLUVIA
116	2018	Nov	25	10.59	1.6	67.79	548.37	2592	4.8	4.5	LLUVIA
117	2018	Nov	26	9.48	1.2	76.5	550.74	2659	14.4	2.7	LLUVIA
118	2018	Dic	3	15.38	0	17.75	548.53	4613	16	4.9	SIN LLUVIA
119	2018	Dic	4	11.93	0	19.67	548.8	8263	16	5.4	SIN LLUVIA
120	2018	Dic	5	12.5	0	21.71	548.43	8037	16	4.9	SIN LLUVIA
121	2018	Dic	6	11.12	0	41.75	549.72	5568	14.7	4.5	SIN LLUVIA
122	2018	Dic	7	8.84	2.2	67.71	552.55	2922	7	3.6	LLUVIA
123	2018	Dic	8	9.39	4	68.92	554.24	4321	14.1	4.5	LLUVIA
124	2018	Dic	9	11.07	0	51.67	552.94	7423	14.9	4.5	SIN LLUVIA
125	2018	Dic	10	10.84	0	50.38	552.63	6074	13.3	6.3	SIN LLUVIA
126	2018	Dic	11	11.09	0	51.25	553.23	4958	10.8	4.9	SIN LLUVIA
127	2018	Dic	12	11.86	0	46.58	552.08	6534	15.5	5.4	SIN LLUVIA
128	2018	Dic	13	11.88	0	45.04	551.1	6582	2.5	4.5	SIN LLUVIA
129	2018	Dic	14	12.65	1.02	49.29	549.79	7690	16	5.8	LLUVIA
130	2018	Dic	15	9.53	0.51	66.71	551.3	4135	11.5	5.4	LLUVIA
131	2018	Dic	16	10.26	7.1	63.96	551.9	3755	11.3	3.6	LLUVIA
132	2018	Dic	17	8.8	5.07	78.25	551.84	3774	15.1	6.7	LLUVIA

UNIVERSIDAD NACIONAL DE HUANCAMELICA - FACULTAD DE CIENCIAS DE INGENIERIA - BASE DE DATOS  
 UNIDAD DE POSGRADO - DOCTORADO EN CIENCIAS AMBIENTALES - M.Sc. Jorge Luis HUERE PEÑA

ORD	año	mes	dia	Temp Out	Rain	Out Hum	Bar mmHg	Solar Rad	UV Index máx	Speed máx	Rains
133	2018	Dic	18	9.02	1.01	72.71	552.57	5135	15	5.8	LLUVIA
134	2018	Dic	19	10.37	0	63.25	552.11	4231	9	3.6	SIN LLUVIA
135	2018	Dic	20	9.5	1.77	69.54	550.5	4727	13.6	4.9	LLUVIA
136	2018	Dic	21	9.08	0	67.21	551.82	4374	12.4	4.9	SIN LLUVIA
137	2018	Dic	22	9.87	2.54	58.54	552.02	7249	16	6.7	LLUVIA
138	2018	Dic	23	10.46	0.51	62.54	550.73	4674	15.5	4.9	LLUVIA
139	2018	Dic	24	10.43	1.52	55.17	550.57	7259	16	5.8	LLUVIA
140	2018	Dic	25	11.11	0.25	58.75	550.31	6478	13.2	4.5	LLUVIA
141	2018	Dic	26	11.16	0	61.33	549.03	3973	12.1	4.5	SIN LLUVIA
142	2018	Dic	27	10.63	0	61.04	549.94	4722	12.7	4.5	SIN LLUVIA
143	2018	Dic	28	10.65	0	61.25	550.2	5291	10.5	6.7	SIN LLUVIA
144	2018	Dic	29	9.02	4.57	62.25	551.05	5153	14.5	4	LLUVIA
145	2018	Dic	30	10.24	0	59.75	550.89	7443	16	4.5	SIN LLUVIA
146	2018	Dic	31	8.76	7.62	65.67	550.61	5609	0	4.9	LLUVIA
147	2019	Ene	1	10.48	0	62.96	551.28	4892	13.1	3.6	SIN LLUVIA
148	2019	Ene	2	11.82	0	57.04	549.79	4527	11.5	3.1	SIN LLUVIA
149	2019	Ene	3	11.3	4.06	59.17	550	5158	14.2	4	LLUVIA
150	2019	Ene	4	10.65	1.26	65.92	550.59	5937	13.8	4.9	LLUVIA
151	2019	Ene	5	10.43	0.25	65.71	551.74	5956	15.7	5.8	LLUVIA
152	2019	Ene	6	12.96	1.01	50.67	549.99	7550	16	3.6	LLUVIA
153	2019	Ene	7	13.82	0	36.17	548.06	7838	16	3.1	SIN LLUVIA
154	2019	Ene	8	11.94	5.84	50.71	548.46	6489	14.7	4.5	LLUVIA
155	2019	Ene	9	11.82	0	59.08	549.37	7822	15.4	4.9	SIN LLUVIA
156	2019	Ene	10	12.55	5.07	51.38	547.25	7770	16	4.5	LLUVIA
157	2019	Ene	11	10.45	4.31	64.88	548.79	7331	15.5	5.4	LLUVIA
158	2019	Ene	12	9.05	6.35	73.92	550.99	3904	12.8	4	LLUVIA
159	2019	Ene	13	8.7	32	76.33	551.81	4564	14.5	2.2	LLUVIA
160	2019	Ene	14	9.04	3.04	78.71	551.75	3839	14.1	3.6	LLUVIA
161	2019	Ene	15	10.14	5.83	72.08	551.29	3496	10.4	3.6	LLUVIA
162	2019	Ene	16	9.45	7.35	70.67	552.51	3664	7.3	4.5	LLUVIA
163	2019	Ene	17	10.05	6.1	72.25	551.85	3396	8.2	3.1	LLUVIA
164	2019	Ene	18	7.09	12.7	81.92	552.86	2400	5.9	3.6	LLUVIA
165	2019	Ene	19	8.93	6.34	76.46	554.02	4269	9.8	3.1	LLUVIA
166	2019	Ene	20	10.6	0	66.13	552.43	4966	11.1	3.1	SIN LLUVIA
167	2019	Ene	21	8.55	15.48	78.67	552.38	3238	6.7	3.6	LLUVIA
168	2019	Ene	22	9.9	5.58	73.83	552.51	3040	7.8	2.7	LLUVIA
169	2019	Ene	23	9.38	6.85	77.21	552.31	2538	10	4	LLUVIA
170	2019	Ene	24	9.29	5.07	78.71	552.96	2891	8.8	3.1	LLUVIA
171	2019	Ene	25	8.68	9.13	81.04	553.14	4762	14.8	3.6	LLUVIA
172	2019	Ene	26	9.5	3.31	73.29	553.21	4437	13	7.2	LLUVIA
173	2019	Ene	27	9.13	5.08	78.46	552.16	3417	10.6	3.1	LLUVIA
174	2019	Ene	28	10.26	0.25	71.04	551.24	5113	15.4	5.4	LLUVIA
175	2019	Ene	29	9.08	14.97	80.63	551.12	4141	12.6	4	LLUVIA
176	2019	Ene	30	8.29	6.08	84.75	551.78	2761	8.5	3.1	LLUVIA

UNIVERSIDAD NACIONAL DE HUANCVELICA - FACULTAD DE CIENCIAS DE INGENIERIA - BASE DE DATOS  
 UNIDAD DE POSGRADO - DOCTORADO EN CIENCIAS AMBIENTALES- M.Sc. Jorge Luis HUERE PEÑA

ORD	año	mes	día	Temp Out	Rain	Out Hum	Bar mmHg	Solar Rad	UV Index máx	Speed máx	Rains
177	2019	Ene	31	8.6	6.6	83.08	552.31	2779	11.5	3.6	LLUVIA
178	2019	Feb	1	8.52	6.1	77.25	553.04	4427	15.1	3.6	LLUVIA
179	2019	Feb	2	9.5	4.57	75.75	552	5437	15	4.9	LLUVIA
180	2019	Feb	3	8.95	5.07	83.96	551.49	2513	9.1	4	LLUVIA
181	2019	Feb	4	9.07	8.12	80.29	552.68	3344	12.9	3.1	LLUVIA
182	2019	Feb	5	10.88	0.25	66.38	551.96	4432	11.7	3.1	LLUVIA
183	2019	Feb	6	9.11	25.39	74.17	551.24	5077	14.8	4.9	LLUVIA
184	2019	Feb	7	7.45	23.36	77.21	552.97	4994	15	4.5	LLUVIA
185	2019	Feb	8	8.27	8.62	78.46	552.23	4237	13.6	4.5	LLUVIA
186	2019	Feb	9	8.84	2.02	75.42	552.26	3786	11.8	4	LLUVIA
187	2019	Feb	10	9.11	7.11	70.75	551.7	4408	15.3	4.9	LLUVIA
188	2019	Feb	11	8.56	11.43	77.67	552.23	4037	11.3	3.6	LLUVIA
189	2019	Feb	12	7.81	3.04	83.75	554.2	2621	7.6	3.1	LLUVIA
190	2019	Feb	13	9.73	14.23	74.88	553.58	6028	15.9	6.3	LLUVIA
191	2019	Feb	14	9.15	1.5	77.58	552.35	4588	14.5	4.5	LLUVIA
192	2019	Feb	15	8.88	3.04	76.17	552.55	2880	13.6	3.6	LLUVIA
193	2019	Feb	16	9.25	4.81	83.43	552.21	2473	8.3	4	LLUVIA
194	2019	Feb	17	10.24	7.63	80	551.06	4146	10.9	4	LLUVIA
195	2019	Feb	18	9.18	4.31	83.33	551.72	5250	14.9	4	LLUVIA
196	2019	Feb	19	10.31	11.43	72.75	551.77	5156	12.1	4.9	LLUVIA
197	2019	Feb	20	8.04	5.32	77.6	551.86	2000	9.6	3.1	LLUVIA
198	2019	Feb	21	9.4	4.31	75.75	552.86	2641	5.6	3.6	LLUVIA
199	2019	Feb	22	11.76	0.25	66.25	551.63	6690	13.7	4.9	LLUVIA
200	2019	Feb	23	10.51	0	68.83	550.44	3489	9.2	4.5	SIN LLUVIA
201	2019	Feb	24	10.95	1.77	65.13	549.94	5583	14.1	4.5	LLUVIA
202	2019	Feb	25	10.75	1.52	62.25	549.96	5806	14.9	4	LLUVIA
203	2019	Feb	26	9.54	1.76	73.71	550.24	4471	15.6	4.5	LLUVIA
204	2019	Feb	27	10.01	0	72.13	550.91	3148	11.7	6.7	SIN LLUVIA
205	2019	Feb	28	10.51	0	64.79	551.1	4910	11.5	4.5	SIN LLUVIA
206	2019	Mar	1	10.6	1.78	67.29	551.04	4436	11.9	4	LLUVIA
207	2019	Mar	2	10.09	9.41	70.13	551.47	3956	10.2	4.9	LLUVIA
208	2019	Mar	3	10.54	0	63.96	551.38	5484	11.8	5.4	SIN LLUVIA
209	2019	Mar	4	9.66	1.52	68.21	551.48	3911	11.2	4	LLUVIA
210	2019	Mar	5	9.76	6.09	68.46	552.32	4293	13.3	3.6	LLUVIA
211	2019	Mar	6	8.9	7.36	76.21	553.08	3704	11.3	4	LLUVIA
212	2019	Mar	7	10.25	3.55	74.46	552.12	4209	14.2	3.1	LLUVIA
213	2019	Mar	8	10.57	1.78	69.08	550.75	3355	10	3.6	LLUVIA
214	2019	Mar	9	9.42	5.82	76.08	551.35	3208	8.1	3.6	LLUVIA
215	2019	Mar	10	9.5	15.99	76.13	552.33	3617	8.7	2.2	LLUVIA
216	2019	Mar	11	9.82	10.66	74.25	552.01	3604	12	3.1	LLUVIA
217	2019	Mar	12	9.07	14.98	78.83	551.12	3423	11.5	2.7	LLUVIA
218	2019	Mar	13	9.68	1.52	69.25	551.6	6374	15.7	4	LLUVIA
219	2019	Mar	14	10.02	1.01	71.96	551.22	3234	11.3	3.6	LLUVIA
220	2019	Mar	15	8.08	8.87	80.92	552.35	2869	9.5	3.6	LLUVIA

UNIVERSIDAD NACIONAL DE HUANCAMELICA - FACULTAD DE CIENCIAS DE INGENIERIA - BASE DE DATOS  
 UNIDAD DE POSGRADO - DOCTORADO EN CIENCIAS AMBIENTALES- M.Sc. Jorge Luis HUERE PEÑA

ORD	año	mes	día	Temp Out	Rain	Out Hum	Bar mmHg	Solar Rad	UV Index máx	Speed máx	Rains
221	2019	Mar	16	9.88	2.79	75.5	552.04	3782	13.8	3.1	LLUVIA
222	2019	Mar	17	9.92	7.87	76.75	550.75	3022	13.1	3.6	LLUVIA
223	2019	Mar	18	9.31	8.38	74.71	551.7	3631	11.6	4.9	LLUVIA
224	2019	Mar	19	10.28	8.89	69.38	551.63	6958	16	4	LLUVIA
225	2019	Mar	20	9.48	0	73.25	551.16	2819	7.8	3.1	SIN LLUVIA
226	2019	Mar	21	9.19	9.13	77.33	551.83	3736	12.3	3.1	LLUVIA
227	2019	Mar	22	9.69	15.22	77.58	551.88	4612	15	4	LLUVIA
228	2019	Mar	23	8.4	17.27	83.21	552.92	2735	8.6	3.1	LLUVIA
229	2019	Mar	24	9.16	22.85	78.92	553.08	3644	12.9	3.6	LLUVIA
230	2019	Mar	25	8.95	5.07	81.08	552.16	3576	12.9	3.6	LLUVIA
231	2019	Mar	26	9.88	0.5	70	551.43	4765	11.9	4	LLUVIA
232	2019	Mar	27	10.55	2.28	71.5	551.05	3065	7.9	4	LLUVIA
233	2019	Mar	28	8.9	5.56	78.75	552.53	2628	5.7	3.1	LLUVIA
234	2019	Mar	29	9.71	0.5	72.13	553.01	3498	11.3	3.1	LLUVIA
235	2019	Mar	30	9.46	0	69.04	553.13	3374	6	3.6	SIN LLUVIA
236	2019	Mar	31	11.2	0	64.78	552.22	4192	10.7	3.1	SIN LLUVIA
237	2019	Abril	1	11.25	1.77	66	550.52	4424	12.9	4	LLUVIA
238	2019	Abril	2	9.71	1.01	76	550.94	3372	12.8	4	LLUVIA
239	2019	Abril	3	10.9	0	59.83	551.65	4896	15.1	4.9	SIN LLUVIA
240	2019	Abril	4	11.25	0	38.5	550.92	7002	15.9	4.9	SIN LLUVIA
241	2019	Abril	5	9.73	0	39.92	552	6838	15.9	3.1	SIN LLUVIA
242	2019	Abril	6	9.49	1.78	59.21	552.02	6093	12.8	4.5	LLUVIA
243	2019	Abril	7	9.19	0	64.54	553.18	3168	9.2	2.7	SIN LLUVIA
244	2019	Abril	8	9.97	0	63.29	552.91	3081	12.3	2.2	SIN LLUVIA
245	2019	Abril	9	10.48	1.27	57.46	552.02	5981	15.9	5.4	LLUVIA
246	2019	Abril	10	9.57	3.81	59.46	551.48	5484	16	5.8	LLUVIA
247	2019	Abril	11	10.42	0	53.29	552.16	5944	14.6	3.1	SIN LLUVIA
248	2019	Abril	12	9.36	2.29	61.33	551.67	3978	13.2	3.6	LLUVIA
249	2019	Abril	13	9.3	6.86	60.83	552.17	6186	15.4	4	LLUVIA
250	2019	Abril	14	8.7	4.57	70.67	552.69	3761	11.8	4	LLUVIA
251	2019	Abril	15	9.79	0	64.58	553.06	3104	10.2	4	SIN LLUVIA
252	2019	Abril	16	9.2	0	66.29	553.33	3064	9.1	4	SIN LLUVIA
253	2019	Abril	17	9	3.04	74.17	552.67	2831	8.3	4	LLUVIA
254	2019	Abril	18	9.34	0.75	72.08	552.52	2624	5.8	2.2	LLUVIA
255	2019	Abril	19	10.75	0.25	66.46	551.94	3738	8.9	3.6	LLUVIA
256	2019	Abril	20	10.1	0.25	67.25	552.19	2884	6.7	4	LLUVIA
257	2019	Abril	21	10.88	0	57.71	552.27	4886	14.5	3.6	SIN LLUVIA
258	2019	Abril	22	9.3	1.52	70	550.95	3486	10.2	5.4	LLUVIA
259	2019	Abril	23	8.28	10.92	73.21	552.93	4384	12.2	4	LLUVIA
260	2019	Abril	24	8.9	2.03	78	552.89	1639	4.3	2.2	LLUVIA
261	2019	Abril	25	9.8	1.51	75.54	553.36	2218	4.6	3.1	LLUVIA
262	2019	Abril	26	10.67	0	65.83	553.21	3118	7	4	SIN LLUVIA
263	2019	Abril	27	10.75	0	59.29	551.65	5008	12.2	3.6	SIN LLUVIA
264	2019	Abril	28	9.35	0.51	63.79	552.12	3023	7.8	3.1	LLUVIA

UNIVERSIDAD NACIONAL DE HUANCAMELICA - FACULTAD DE CIENCIAS DE INGENIERIA - BASE DE DATOS  
 UNIDAD DE POSGRADO - DOCTORADO EN CIENCIAS AMBIENTALES - M.Sc. Jorge Luis HUERE PEÑA

ORD	año	mes	día	Temp Out	Rain	Out Hum	Bar mmHg	Solar Rad	UV Index máx	Speed máx	Rains
265	2019	Abril	29	9.61	0.5	66.71	552.32	3438	10.2	4.5	LLUVIA
266	2019	Abril	30	8.92	1.77	70.04	552.39	3326	7.8	4.5	LLUVIA
267	2019	May	1	8.43	0	71.92	553.68	5026	10.9	4.5	SIN LLUVIA
268	2019	May	2	9.62	0.25	65.19	552.97	4377	14.4	4	LLUVIA
269	2019	May	3	10.78	0.51	57.71	562.47	2443	9.2	3.6	LLUVIA
270	2019	May	4	9.56	0	57.83	569.41	3275	7.8	2.7	SIN LLUVIA
271	2019	May	5	8.31	4.82	71.75	570.86	2411	6.4	3.6	LLUVIA
272	2019	May	6	9.78	0.25	68.46	571.77	2556	7.6	3.1	LLUVIA
273	2019	May	7	8.66	0	61.71	572.44	3117	5.9	3.6	SIN LLUVIA
274	2019	May	8	8.65	0.25	63.17	571.32	2912	8.7	3.1	LLUVIA
275	2019	May	9	9.72	5.08	65.79	570.12	3391	8.3	3.1	LLUVIA
276	2019	May	10	10.41	0.25	54.92	569.66	5397	13.4	3.1	LLUVIA
277	2019	May	11	11.51	0.76	54.57	567.24	1790	8.5	4	LLUVIA
278	2019	May	12	10.36	0.25	60.92	569.22	3206	11.2	3.6	LLUVIA
279	2019	May	13	9.03	5.33	76.5	570.38	3545	10	3.6	LLUVIA
280	2019	May	14	8.77	4.82	79	571.53	2880	6.2	2.7	LLUVIA
281	2019	May	15	9.27	0.25	61.29	571.8	5030	13.6	3.6	LLUVIA
282	2019	May	16	10	0	46.88	570.94	5842	13	2.7	SIN LLUVIA
283	2019	May	17	9.84	0	45.79	569.83	5848	12.8	2.2	SIN LLUVIA
284	2019	May	18	9.1	0	45.21	569.96	5306	12.2	3.1	SIN LLUVIA
285	2019	May	19	10.18	0	40.04	570.38	4923	11.6	4.9	SIN LLUVIA
286	2019	May	20	9.9	0	34.67	570.01	4861	10.5	4	SIN LLUVIA
287	2019	May	21	9.01	0	35.71	570.67	5526	11.8	3.6	SIN LLUVIA
288	2019	May	22	8.34	0	55.08	571.06	3378	8.3	3.6	SIN LLUVIA
289	2019	May	23	8.59	0	53.46	571.86	4813	11.8	4.5	SIN LLUVIA
290	2019	May	24	8.7	0	45.33	571.47	5726	11.8	3.6	SIN LLUVIA
291	2019	May	25	8.82	10.41	59.71	570.11	4466	10.8	4	LLUVIA
292	2019	May	26	7.98	0.51	56.75	571.84	3347	8.8	3.1	LLUVIA
293	2019	May	27	8.96	0.76	46.75	571.72	5572	11.4	4.5	LLUVIA
294	2019	May	28	8.54	0	48.08	571.28	4512	11.2	2.2	SIN LLUVIA
295	2019	May	29	9.12	0	46.54	571.98	5154	11.4	3.1	SIN LLUVIA
296	2019	May	30	9.65	0	45.17	571.57	5563	12.2	3.1	SIN LLUVIA
297	2019	May	31	9.83	0	39.42	570.84	5534	12.2	4.5	SIN LLUVIA
298	2019	Jun	1	9.51	0	38.5	570.72	5460	11.7	3.6	SIN LLUVIA
299	2019	Jun	2	9.22	0	44.71	571.02	5329	11.8	3.6	SIN LLUVIA
300	2019	Jun	3	8.41	0	55.38	572.52	4488	9.4	4	SIN LLUVIA
301	2019	Jun	4	8.7	0	53.63	573.35	3967	9.4	3.1	SIN LLUVIA
302	2019	Jun	5	8.14	0	51.13	572.97	4931	10.6	3.6	SIN LLUVIA
303	2019	Jun	6	7.22	0	34.04	573.62	5627	11.3	3.1	SIN LLUVIA
304	2019	Jun	7	7.55	0	28.54	573.24	5602	11.4	3.1	SIN LLUVIA
305	2019	Jun	8	8.17	0	29.5	572.96	5501	11.1	4	SIN LLUVIA
306	2019	Jun	9	8.09	0	38.5	573.14	5382	10.8	4	SIN LLUVIA
307	2019	Jun	10	9.63	0	44.32	571.54	4810	10.8	3.6	SIN LLUVIA
308	2019	Jun	11	9.68	0	46.58	571.16	5161	8.9	4	SIN LLUVIA

UNIVERSIDAD NACIONAL DE HUANCAMELICA - FACULTAD DE CIENCIAS DE INGENIERIA - BASE DE DATOS  
 UNIDAD DE POSGRADO - DOCTORADO EN CIENCIAS AMBIENTALES- M.Sc. Jorge Luis HUERE PEÑA

ORD	año	mes	día	Temp Out	Rain	Out Hum	Bar mmHg	Solar Rad	UV Index máx	Speed máx	Rains
309	2019	Jun	12	9.22	0	44	571.45	5139	10.9	3.1	SIN LLUVIA
310	2019	Jun	13	8.89	0	36.08	572.39	4364	9.9	3.6	SIN LLUVIA
311	2019	Jun	14	9.13	0	36.58	572.43	5234	10.8	3.6	SIN LLUVIA
312	2019	Jun	15	8.28	0	48.79	571.32	4744	10.7	3.6	SIN LLUVIA
313	2019	Jun	16	7.95	0.25	58.92	572.07	2714	6.6	4.5	LLUVIA
314	2019	Jun	17	9.22	0	57.75	572.34	3374	9	3.1	SIN LLUVIA
315	2019	Jun	18	9.67	0	51.38	571.81	5223	9.7	3.6	SIN LLUVIA
316	2019	Jun	19	8.75	0	49.96	572.01	5679	10.3	4.5	SIN LLUVIA
317	2019	Jun	20	6.55	0	56.38	574.08	2424	5	2.7	SIN LLUVIA
318	2019	Jun	21	6.88	0	42.54	576.26	5464	10.6	4	SIN LLUVIA
319	2019	Jun	22	6.67	0	28.71	576.04	5510	10.9	4.9	SIN LLUVIA
320	2019	Jun	23	6.43	0	21.04	575.71	5516	10.7	4	SIN LLUVIA
321	2019	Jun	24	7.48	0	22.25	574.12	5499	10.8	4	SIN LLUVIA
322	2019	Jun	25	8.4	0	23.17	571.72	5482	10.7	4.9	SIN LLUVIA
323	2019	Jun	26	7.41	0	35.58	571.31	5389	9.7	4.5	SIN LLUVIA
324	2019	Jun	27	9.09	0	47.08	570.43	5205	10.6	3.1	SIN LLUVIA
325	2019	Jun	28	9.54	0	57.17	569.26	2843	7.8	2.7	SIN LLUVIA
326	2019	Jun	29	9.46	0	58.04	570.03	3098	9.6	4.5	SIN LLUVIA
327	2019	Jun	30	9.63	0	54.21	570.33	2152	4.3	2.7	SIN LLUVIA
328	2019	Jul	1	8.42	0	45.79	572.22	5793	11	3.1	SIN LLUVIA
329	2019	Jul	2	8.45	0	42.08	571.66	5220	10.8	3.6	SIN LLUVIA
330	2019	Jul	3	10.38	0	21.5	569.77	5532	11	3.6	SIN LLUVIA
331	2019	Jul	4	9.55	0	27.08	569.68	5434	10.5	4.5	SIN LLUVIA
332	2019	Jul	5	8.77	0	42.38	571.31	5628	10.4	4	SIN LLUVIA
333	2019	Jul	6	7.23	0	50.63	573.38	3165	7.8	4	SIN LLUVIA
334	2019	Jul	7	8.25	0	50	573.08	4408	11.1	3.1	SIN LLUVIA
335	2019	Jul	8	8.41	0	46.88	572.39	5312	10.7	3.6	SIN LLUVIA
336	2019	Jul	9	7.68	0	38.58	573.54	5441	11.1	4	SIN LLUVIA
337	2019	Jul	10	7.13	0	30.88	573.93	5597	11	4	SIN LLUVIA
338	2019	Jul	11	6.98	0	17.58	573.55	5643	10.9	4	SIN LLUVIA
339	2019	Jul	12	8.67	0	23.54	571.8	5542	10.7	3.6	SIN LLUVIA
340	2019	Jul	13	8.08	0	24.92	571.43	5514	10.5	4	SIN LLUVIA
341	2019	Jul	14	7.72	0	39.83	571.84	5452	10.8	4	SIN LLUVIA
342	2019	Jul	15	8.2	0	44.04	571.48	5034	10.9	3.6	SIN LLUVIA
343	2019	Jul	16	8.55	0	49.67	571.51	3227	6.5	3.6	SIN LLUVIA
344	2019	Jul	17	7.18	3.29	66.75	571.4	2340	5.8	4	LLUVIA
345	2019	Jul	18	7.95	1.52	62.25	573.12	3194	7	2.7	LLUVIA
346	2019	Jul	19	8.24	0	57.96	571.27	4130	10.2	3.1	SIN LLUVIA
347	2019	Jul	20	8.15	0	51.75	572.18	5446	10.8	3.6	SIN LLUVIA
348	2019	Jul	21	8.64	0	50.46	571.09	5305	12.3	4.5	SIN LLUVIA
349	2019	Jul	22	9.48	0	45.92	570.29	4944	10.5	4.9	SIN LLUVIA
350	2019	Jul	23	8.08	0.5	54.83	570.3	3499	8.4	3.6	LLUVIA
351	2019	Jul	24	8.39	0	42.17	572.08	4438	11.7	4.5	SIN LLUVIA
352	2019	Jul	25	7.69	0	26.79	572.66	5723	12.2	3.6	SIN LLUVIA

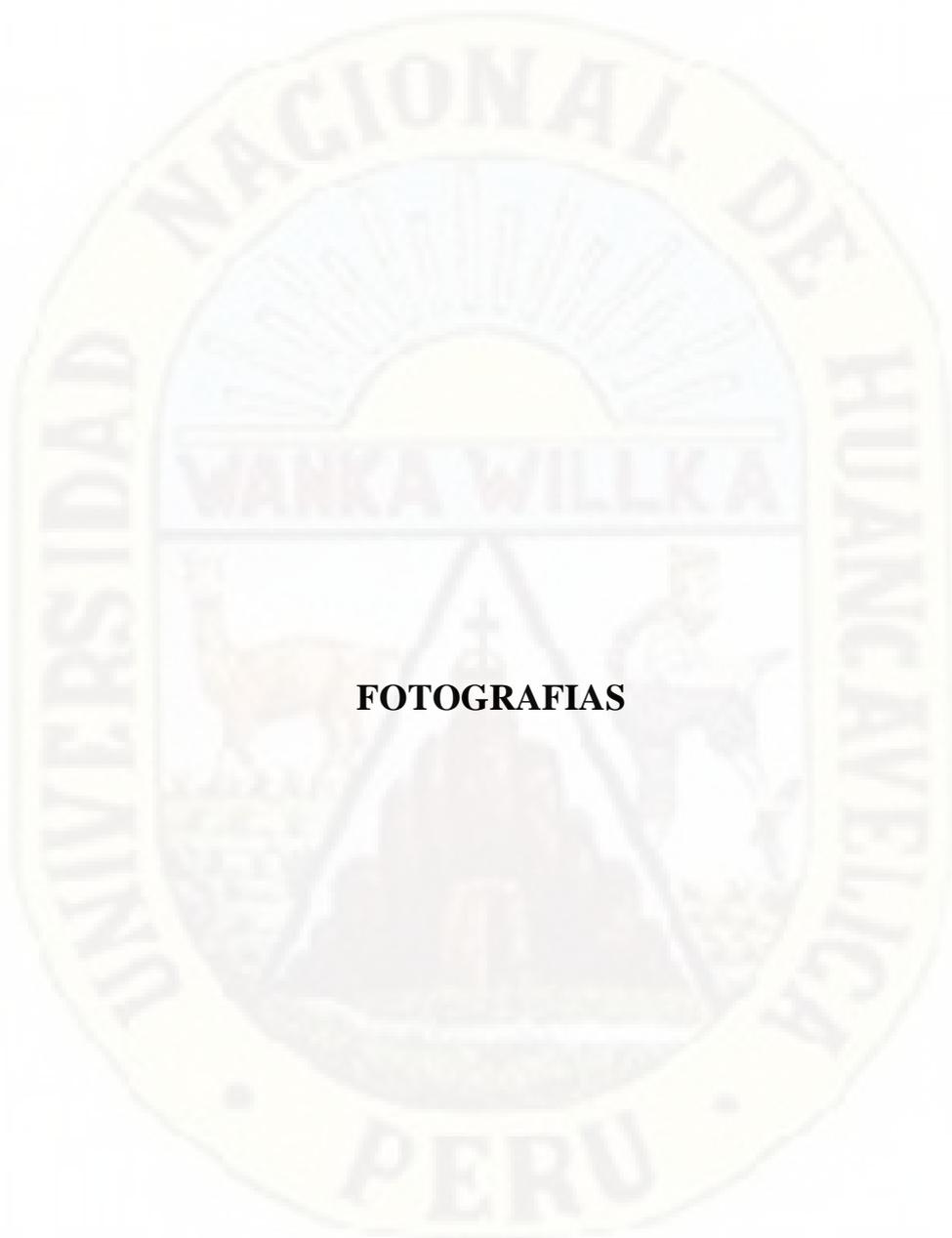
UNIVERSIDAD NACIONAL DE HUANCAMELICA - FACULTAD DE CIENCIAS DE INGENIERIA - BASE DE DATOS  
 UNIDAD DE POSGRADO - DOCTORADO EN CIENCIAS AMBIENTALES - M.Sc. Jorge Luis HUERE PEÑA

ORD	año	mes	día	Temp Out	Rain	Out Hum	Bar mmHg	Solar Rad	UV Index máx	Speed máx	Rains
353	2019	Jul	26	7.43	0	34.08	573.28	5691	12.2	4	SIN LLUVIA
354	2019	Jul	27	8.44	0	50.33	572.42	3743	7.7	3.6	SIN LLUVIA
355	2019	Jul	28	9.12	0	53.33	571.51	3005	7.7	4.5	SIN LLUVIA
356	2019	Jul	29	9.25	0	40.79	571.5	4503	12.6	3.6	SIN LLUVIA
357	2019	Jul	30	8.45	0	31.88	572.67	5772	12	2.7	SIN LLUVIA
358	2019	Jul	31	8.06	0	33.17	572.83	5705	11.9	4	SIN LLUVIA





**ARTÍCULO CIENTÍFICO**



**FOTOGRAFIAS**

**Figura 32**

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



**Figura 33**

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



**Figura 34**

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



**Figura 35**

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

