

REGLAS DE CLASIFICACIÓN DE ESTRÉS TÉRMICO POR CALOR A PARTIR DE ÁRBOLES DE DECISIÓN PARA LA CIUDAD DE TARIJA

CLASSIFICATION RULES FOR THERMAL HEAT STRESS FROM DECISION TREES FOR THE CITY OF TARIJA

Nieves Soledad Vásquez Perales¹

¹ Doctorando. Universidad Autónoma Juan Misael Saracho, Doctorado en Ciencias. Tarija, Bolivia.

Correo electrónico: vasquez.perales.nieves@gmail.com

RESUMEN

El objetivo del presente trabajo de investigación es determinar las reglas de clasificación de estrés térmico por calor a partir de árboles de decisión para la ciudad de Tarija. Como insumo principal, se hace uso de una base de datos meteorológica que contiene los datos de entrada o instancias que corresponden a los registros diarios de la estación meteorológica Tarija Aeropuerto comprendidos entre los años 2017 y 2021 obtenidos del Servicio Nacional de Meteorología e Hidrología. La determinación del nivel de riesgo que experimentan las personas que desarrollan actividades al aire libre se basa en la herramienta propuesta por el Servicio Meteorológico Nacional de los Estados Unidos para evaluar el potencial del estrés por calor denominada índice de calor, cuya medida se basa en la temperatura del aire y la humedad relativa. La técnica de minería de datos utilizada para la construcción del modelo predictivo son los árboles de decisión aplicando los algoritmos K-medias y C4.5, cuyo resultado arroja un total de 12 reglas de clasificación de estrés térmico por calor, donde 0.61% de las instancias se clasifican con nivel de riesgo de peligro, 8.28% de precaución extrema y 33.61% de precaución. Ninguna instancia se clasifica con peligro extremo. El 57.50% no aplica para la existencia de nivel de riesgo. Finalmente, el árbol de decisión obtenido en el marco del objeto de estudio, pone en evidencia de que las principales variables meteorológicas que inciden en el cálculo del nivel de riesgo por estrés térmico son la temperatura y humedad relativa.

PALABRAS CLAVE

Reglas de clasificación, estrés térmico por calor, datos meteorológicos, árboles de decisión.

ABSTRACT

The objective of this research work is to determine the classification rules for thermal heat stress from decision trees for the city of Tarija. As the main input, a meteorological database is used that contains the input

data or instances that correspond to the daily records of the Tarija Aeropuerto meteorological station between the years 2017 and 2021 obtained from the National Meteorology and Hydrology Service. The determination of the level of risk experienced by people who carry out outdoor activities is based on the tool proposed by the United States National Weather Service to assess the potential for heat stress called the heat index, whose measurement is based on air temperature and relative humidity. The data mining technique used to build the predictive model is decision trees applying the K-means and C4.5 algorithms, the result of which yields a total of 12 thermal heat stress classification rules, where 0.61% of the instances are classified with danger risk level, 8.28% extreme caution and 33.61% caution. No instance is classified with extreme danger. 57.50% does not apply for the existence of a level of risk. Finally, the decision tree obtained within the framework of the object of study, shows that the main meteorological variables that affect the calculation of the level of risk due to thermal stress are temperature and relative humidity.

KEYWORDS

Classification rules, thermal heat stress, meteorological data, decision trees.

INTRODUCCIÓN

Cuando se tiene niveles altos de temperatura en el ambiente, puede ser agobiante para quienes desempeñan actividades laborales al aire libre; este hecho se incrementa cuando la humedad relativa es alta o no corre el aire.

Según el Instituto Nacional de Seguridad e Higiene en el Trabajo español, define al estrés térmico por calor como:

La carga de calor que los trabajadores reciben y acumulan en su cuerpo y que resulta de la interacción entre las condiciones ambientales del lugar donde trabajan, la actividad física que realizan y la

ropa que llevan. Es decir, el estrés térmico por calor no es un efecto patológico que el calor puede originar en los trabajadores, sino la causa de los diversos efectos patológicos que se producen cuando se acumula excesivo calor en el cuerpo. (MTASE, s.f.).

Otra definición de estrés térmico puede indicarse como:

La sensación de malestar que se experimenta cuando la permanencia en un ambiente determinado exige esfuerzos desmesurados a los mecanismos de que dispone el organismo para mantener la temperatura interna, mientras se efectúa el intercambio de agua y demás sustancias. (Barcía, S. et al., 2021).

De acuerdo a 3M (2017), la Tabla 1 incluye las señales de advertencia de estrés térmico.

Tabla 1. Señales de advertencia de estrés térmico por calor

Síntomas de agotamiento por calor	Síntomas de golpe de calor
- Dolores de cabeza, vértigo mareos o desmayos - Debilidad - Piel húmeda - Cambios de humor, tales como irritabilidad o confusión - Malestar estomacal o vómitos	- Confusión mental o pérdida de conciencia - Ataques o convulsiones - Piel seca y caliente sin sudor

Fuente: (3M, 2017)

El Servicio Meteorológico Nacional de los Estados Unidos (MWS) provee múltiples herramientas para evaluar el potencial de estrés térmico por calor, entre las cuales se señalan: índice de calor, temperatura de globo de bulbo húmedo (WBGT) y riesgo por calor (NWS, s.f.a), donde la primera, está basada en el índice denominado temperatura aparente propuesto por Stedman en el año 1979 y lo define como “la temperatura que siente el cuerpo humano cuando la humedad relativa se combina con la temperatura del aire” (NWS, s.f.b), que es la definición que orienta a la presente investigación, siendo las variables meteorológicas mencionadas, las principales para el logro del objetivo del trabajo.

Sin embargo, la complejidad que se da entre la relación del clima y poblaciones está muy ligada a cada ámbito geográfico, lo que implica un estudio específico para cada caso definido (Linares, C. et al., 2017).

En cuanto al ámbito geográfico del área de estudio, el Plan Departamental de Gestión del Riesgo y Adaptación

al Cambio Climático (2014-2018) hace referencia que por la ubicación geográfica del departamento de Tarija entre los 20°50' y 22°50' de latitud Sur y entre los 62°15' y 65°20' de longitud Oeste, tiende a presentar temperaturas altas que son la causa de las sequías ecológicas, presentando también un nivel bajo de precipitaciones (GADT, 2014).

Por otra parte, en este plan también se realiza el análisis de amenazas por pisos ecológicos, cuyo resultado correspondiente a la Cordillera Oriental: Valle Central que comprende los municipios de Tarija, Uriondo, parte de San Lorenzo y parte de Padcaya, se muestra en la Tabla 2.

Tabla 2. Amenazas identificadas en la Cordillera Oriental: Valle Central

Descripción	Amenazas
Es una región de serranías y colinas que rodean los valles de los ríos Guadalquivir, Tolomosa, Santa Ana y Camacho. El valle se encuentra a una altura entre 1750 y 2100 m.s.n.m. Muestra un clima templado semiárido, con temperatura media.	- Granizadas - Heladas - Riadas - Incendios - Sequías - Contaminación

Fuente: (GADT, 2014)

La ciudad de Tarija es la capital del municipio de Tarija y del departamento de Tarija, cuya temperatura media anual es de 17.9 °C con máximas medias anuales de 26.2°C y 9.5°C de mínima media anual, siendo también una característica importante de los últimos años la presencia de temperaturas extremas y sequías, tal como ocurrió en el año 2017 y en enero de 2019, presentándose una máxima extrema de 39.3°C.

En lo referido a las precipitaciones, se presenta una estacionalidad marcada, esto debido a que el mayor porcentaje de presencia de lluvias aparecen entre los meses de octubre y marzo, alcanzando un 94%, siendo la precipitación anual de 596.8 mm.

De acuerdo al estudio realizado por el Instituto de Hidráulica Ambiental de la Universidad de Cantabria que lleva por título “Índice de Vulnerabilidad y Adaptación al Cambio Climático”, realiza la identificación y priorización de medidas de adaptación frente al cambio climático, cuyos resultados obtenidos del análisis realizado a la precipitación y la temperatura media anual en el municipio de Tarija, ponen en evidencia la elevada variabilidad de estas dos variables, en especial con la temperatura, la cual presenta un fuerte gradiente este-oeste (IHCantabria, 2021).

Asimismo, también concluye que, dadas las condiciones del área de estudio, cuya premisa principal es el aumento de la densidad de la población y las

características constructivas de las viviendas, genera un efectivo invernadero localizado, proyectando un incremento de las condiciones de riesgo ante las olas de calor, además de la disminución del confort climático natural para la población. (IHCantabria, 2021).

Entonces, el cambio climático tiene mucha incidencia en el aumento del estrés térmico, no solo en la ciudad de Tarija, sino a nivel mundial. De acuerdo al artículo publicado en el periódico El País, hace referencia que según la Organización Internacional del Trabajo (OIT), el aumento del estrés térmico prevé que traiga consigo 2030 pérdidas de productividad en todo el mundo, equivalentes a 80 millones de puestos de trabajo a tiempo completo; asimismo, indica la estrecha relación con el aumento de calor (El País, 2019).

Por lo tanto, el exceso de calor trae consigo serios efectos en el ámbito laboral constituyéndose en un riesgo para la salud en el trabajo, ocasionando la disminución de la productividad, llegando a ser mortales. Entre los sectores más vulnerables ante el estrés térmico por calor se señalan al agrícola con un 60% y construcción con un 19%; el restante porcentaje lo ocupan los sectores de bienes y servicios medioambientales, recojo de basura, emergencias, trabajos de reparación, transporte, turismo y deportes (El País, 2019).

Son varias las investigaciones orientadas al estudio del estrés térmico desde diferentes ópticas que van desde el análisis del índice de calor como factor de alerta temprana en salud pública y ciudades sostenibles (Salazar y Álvarez, 2020), golpe de calor en trabajos al aire libre (MTMSS, 2018), hasta la determinación de los intervalos de confort y estrés térmico para espacios abiertos (Aquino, M. et al., 2019).

Por otra parte, la aparición de nuevos métodos de análisis inteligente para la extracción útil de conocimiento ha ayudado en el análisis de este tipo de estudios, tal como lo es la minería de datos, la cual recoge ideas y recursos de varias disciplinas como la inteligencia artificial, la estadística, la computación de alto rendimiento, entre otras.

En la actualidad, los procesos de negocio de las organizaciones se basan en sistemas de información, cuyos registros relacionados a la ejecución de sus procesos podrían ser ejecutados por técnicas de minería de datos para aprovechar de mejor manera la información que proveen (Aguirre, H. y Rincón, N., 2015).

La minería de datos es clasificada como parte del llamado proceso KDD - Knowledge Discovery in Databases (descubrimiento de conocimiento en bases de datos), que involucra la búsqueda de modelos y patrones, los cuales, una vez descubiertos han de ser

válidos y potencialmente útiles para la clasificación y descubrimiento de patrones (Valcárcel, V., 2004).

Una de las técnicas utilizadas en la fase de minería de datos son los árboles de decisión, que tuvieron su origen en las ciencias sociales con los trabajos de Sonquist y Morgan (1964) a través del software AID (Automatic Interaction Detection), el cual se constituyó en uno de los primeros métodos de ajuste de datos con base en los árboles de decisión (Bouza, C. y Santiago, A., 2012).

A partir de la observación y el análisis de aspectos concretos, permite desarrollar modelos con el fin de explicar estos datos y establecer una generalización.

Entonces, comprender los beneficios que aporta la utilización de técnicas de minería de datos, permitirá aportar a los estudios ya realizados con relación al estrés térmico y sus implicancias.

Finalmente, el objetivo del presente trabajo de investigación es determinar las reglas de clasificación de estrés térmico por calor a partir de árboles de decisión para la ciudad de Tarija.

MATERIALES Y MÉTODOS

La investigación se circunscribe en el método inductivo-deductivo, porque, por una parte, a través de la técnica de minería de datos de árboles de decisión se establecen reglas de clasificación para generalizar un mapa de posibles resultados por la relación de variables meteorológicas a través de software especializado; por otra parte, se considera también conceptos generalmente aceptados en este campo para determinar explicaciones particulares.

El presente trabajo se enmarca dentro del enfoque mixto, dado que se aplica la recolección de datos basado en el análisis, medición y técnicas de minería de datos, además se investiga sobre la realidad de un acontecimiento, que este caso lo representan las variables meteorológicas que influyen en la determinación del índice de calor; asimismo, se establece el análisis crítico de los resultados encontrados.

En cuanto al alcance, el trabajo es de tipo explicativo, debido a que el fin principal es poder dar respuesta al por qué ocurre ciertos eventos, en este caso es el estrés térmico por calor, además de poder explicar del por qué ciertas variables meteorológicas influyen en la determinación del índice de calor. Entonces, el objetivo no solamente es la descripción de este fenómeno, sino también, determinar las causales para que el mismo suceda.

El diseño es de tipo no experimental longitudinal, dado de que se toma en cuenta como principal insumo a las variables meteorológicas históricas diarias para determinar la relación existente entre estas y su incidencia en la determinación del índice de calor a través del tiempo asociadas a niveles de riesgo y su correspondiente medida de protección.

Base de datos meteorológica

Los datos que se utilizan para la presente investigación son los que corresponden a los valores de las variables meteorológicas comprendidos entre los años 2017 y 2021 de la estación meteorológica Tarija Aeropuerto obtenidos del Servicio Nacional de Meteorología e Hidrología (SENAMHI, s.f.), cuyos datos técnicos se muestran en la Figura 1.

Figura 1. Kardex estación meteorológica Tarija Aeropuerto

KARDEX DE ESTACIÓN	
Código:	A-2164 [Activo]
Nombre:	Tarija Aeropuerto
Tipo:	Meteorológica
Sub tipo:	Convencional
Latitud:	-21.549179
Longitud:	-64.707653
Altitud:	1875
Departamento:	Tarija
Provincia:	Cercado Tarija
Municipio:	Tarija
Fecha creación:	1944-03-30
Dirección:	

Fuente: (SENAMHI, s.f.)

Las variables directamente relacionadas con el cálculo del índice de calor son: temperatura del aire (°C) y humedad relativa (%), y las secundarias para fines de validar las reglas de clasificación son: precipitación (mm), presión atmosférica (hPa), dirección del viento y velocidad del viento (Km/h). Cabe mencionar que se utiliza la temperatura máxima, debido a que el objetivo es la determinación de estrés térmico por calor.

Dentro del conjunto de datos meteorológicos obtenidos, se observa que no se cuenta con la totalidad de valores en la estación meteorológica Tarija Aeropuerto, para lo cual, se aplica la relación de dependencia con otras estaciones meteorológicas cercanas con estado activo, como lo son la estaciones El Tejar y San Andrés, cuyos datos técnicos se muestran en la Figura 2 y Figura 3 respectivamente, con el fin de completar los datos faltantes a través del establecimiento de una ecuación del tipo $y = ax + b$ cuyo coeficiente de determinación R^2 sea lo más cercano a 1 para aceptar la relación.

Figura 2. Kardex estación meteorológica El Tejar

KARDEX DE ESTACIÓN	
Código:	A-2164 [Activo]
Nombre:	El Tejar (Tarija)
Tipo:	Meteorológica
Sub tipo:	Convencional
Latitud:	-21.543056
Longitud:	-64.721111
Altitud:	1859
Departamento:	Tarija
Provincia:	Cercado Tarija
Municipio:	Tarija
Fecha creación:	1969-12-31
Dirección:	

Fuente: (SENAMHI, s.f.)

Figura 3. Kardex estación meteorológica San Andrés

KARDEX DE ESTACIÓN	
Código:	A-2164 [Activo]
Nombre:	San Andres
Tipo:	Meteorológica
Sub tipo:	Convencional
Latitud:	-21.623333
Longitud:	-64.815
Altitud:	1987
Departamento:	Tarija
Provincia:	Cercado Tarija
Municipio:	Tarija
Fecha creación:	1975-06-30
Dirección:	

Fuente: (SENAMHI, s.f.)

Una vez completados los datos faltantes, se tabulan las variables indicadas que se relacionan con el cálculo del índice de calor, adicionando la fecha. Este proceso se realiza en una hoja de cálculo electrónica.

Representación de la información geográfica

Para la representación de la información geográfica del área de estudio, se utiliza la herramienta ArcMap 10.8, que es la aplicación central de ArcGIS, permitiendo esta representación como una colección de capas, además de otros elementos en un mapa (ESRI, s.f.).

Las capas geográficas para la elaboración de mapas se obtienen del portal de GeoBolivia, el cual es una iniciativa de la Vicepresidencia del Estado Plurinacional de Bolivia, que tiene a disposición un catálogo avanzado de ficheros de metadatos e información geográfica de Bolivia, constituyéndose en el vínculo entre instituciones que forman una red cuyo fin es el de compartir sin costo alguno (GeoBolivia, s.f.).

Determinación del índice de calor

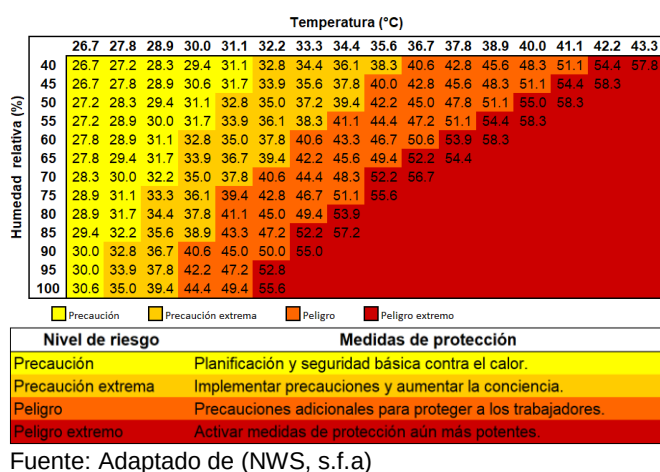
El índice de calor es una medida para establecer la sensación térmica cuando se relaciona la humedad relativa con la temperatura del aire. En el presente trabajo, se utiliza la herramienta de índice de calor propuesto por el Servicio Meteorológico Nacional (NWS, s.f.a).

La Figura 4 muestra la relación de estas dos variables, el nivel de riesgo y las medidas de protección recomendadas, donde se realiza la conversión de la temperatura de grados Fahrenheit (°F), el cual es el dato original de NWS, a grados Celsius (°C) aplicando la fórmula:

$$^{\circ}\text{C} = \frac{^{\circ}\text{F} - 32}{1.8}$$

Asimismo, es importante indicar que estos valores fueron diseñados para condiciones de sombra y viento ligero, entonces, la exposición a la luz solar podría aumentar los valores del índice de calor hasta 8 °C.

Figura 4. Cuadro de índice de calor



Fuente: Adaptado de (NWS, s.f.a)

Posteriormente, por cada uno de los registros diarios tabulados de las variables meteorológicas relacionadas al cálculo del índice de calor, se toma en cuenta la temperatura máxima y la humedad relativa para poder establecer el nivel de riesgo al que corresponden de acuerdo al detalle indicado en la Figura 4, asignando la etiqueta correspondiente (Nivel de riesgo) por cada registro.

Determinación de clusters, árbol de decisión y reglas de clasificación

Se utiliza la herramienta de software libre Weka versión 3.9.6. La Figura 5 muestra la interfaz principal de Weka, a partir del cual se realiza el procesamiento de datos para la determinación de clusters, generación del árbol de decisión y reglas de clasificación del nivel de riesgo de estrés térmico por calor.

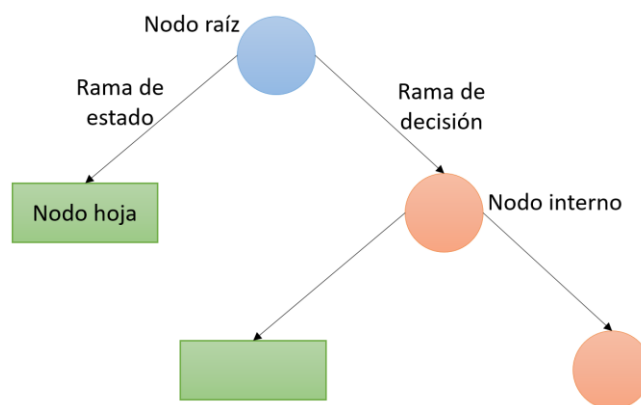
Figura 5. Interfaz principal de Weka



Fuente: Herramienta Weka versión 3.9.6

Para la clasificación de instancias en clusters (asociaciones y estructuras de datos) se aplica el algoritmo K-medias (Pérez, J. et al., 2018). Para la generación del árbol de decisión (Caballero, L. et al., 2015) se aplica el algoritmo C4.5 (J48 en Weka), donde a partir de los resultados que este arroja y por la estructura del árbol de decisión según la Figura 6, se establecen las reglas de clasificación recorriendo los caminos que se van formando desde el nodo raíz, recorriendo por las ramas, nodos internos hasta llegar a los nodos hoja.

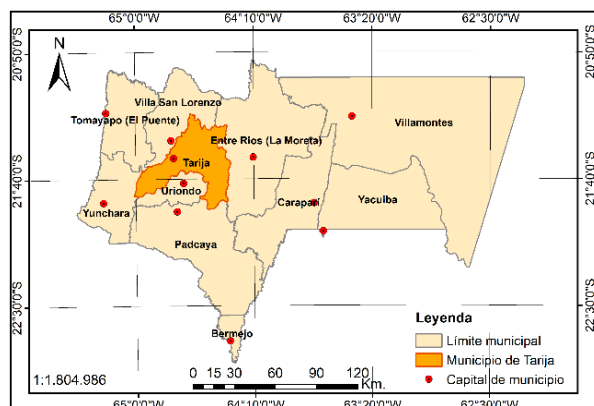
Figura 6. Estructura general de un árbol de decisión



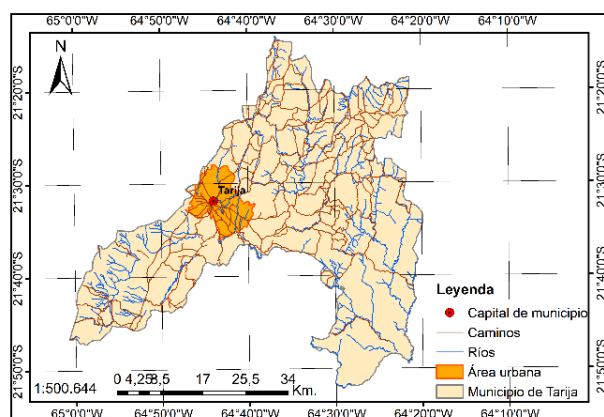
Fuente: Elaboración propia

RESULTADOS

Se identificó geográficamente el área de estudio. La Figura 7 muestra la ubicación geográfica del municipio de Tarija en el departamento de Tarija y la Figura 8 muestra el área geográfica del área urbana del municipio de Tarija.

Figura 7. Ubicación geográfica del municipio de Tarija

Fuente: Elaboración propia

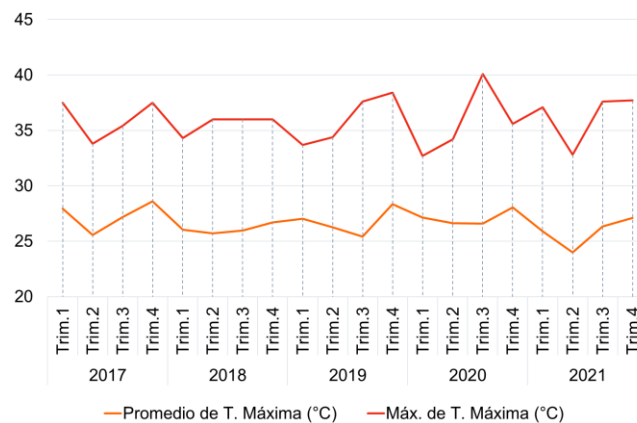
Figura 8. Ubicación geográfica del área urbana del municipio de Tarija

Fuente: Elaboración propia

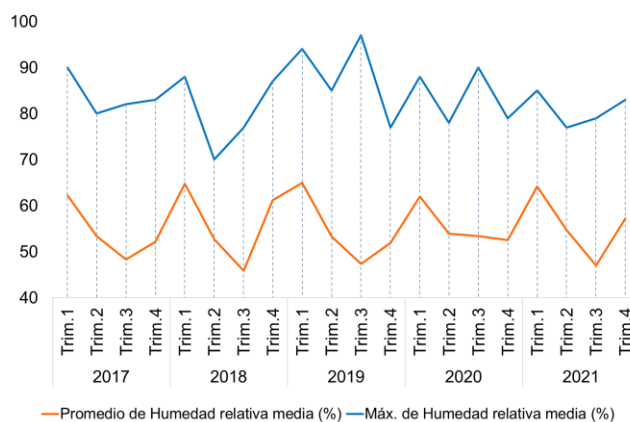
Con el fin de completar los datos faltantes, se relacionaron las variables de temperatura máxima y humedad relativa de las estaciones meteorológicas de San Andrés (x) y Tarija Aeropuerto (y), donde la relación de regresión puede explicar el 98.03% de la variación de la variable dependiente de temperatura máxima y el 94.63% de la variación de la variable dependiente de la humedad relativa. En ambos casos, se acepta la relación para el cálculo de los datos faltantes.

Con respecto a las variables consideradas secundarias en la presente investigación, para el relleno de datos faltantes de precipitación, se relacionaron los datos de las estaciones meteorológicas de El Tejar (x) y Tarija Aeropuerto (y), obteniendo un coeficiente de determinación de 84.28%. En el caso de las variables de presión atmosférica (HPa), dirección del viento y velocidad del viento (Km/h), estas no se relacionaron con otra estación por contar con los datos suficientes para los objetivos del presente trabajo.

A continuación, teniendo los datos completos, la Figura 9 y Figura 10 muestran las tendencias de las variables meteorológicas de temperatura máxima y humedad relativa media entre los años 2017 y 2021 de la estación meteorológica Aeropuerto Tarija.

Figura 9. Tendencia de temperatura máxima en la estación meteorológica Tarija Aeropuerto

Fuente: Elaboración propia

Figura 10. Tendencia de humedad relativa media en la estación meteorológica Tarija Aeropuerto

Fuente: Elaboración propia

De acuerdo a lo establecido por el Servicio Meteorológico Nacional (NWS) a través de su cuadro de índice de calor detallado en la Figura 4, se realizó la relación de la temperatura máxima y la humedad relativa, obteniendo clasificaciones diarias del nivel de riesgo. La Tabla 3 muestra una parte del conjunto de variables asociadas al nivel de riesgo asociado.

Tabla 3. Nivel de riesgo diario

Fecha	Temp. máxima (°C)	Hum. relativa media (%)	Precipitación (mm)	Pres. atmosf. media (hPa)	Dir. del viento	Vel. media viento (Km/h)	Nivel de riesgo
01/01/2017	33.5	52.0	0.3	813	S	6.0	Prec_Ext
02/01/2017	31.6	55.0	0.0	815	SE	10.0	Prec_Ext
03/01/2017	33.0	43.0	0.0	811	SSW	5.0	Prec_Ext
04/01/2017	33.2	48.0	4.0	809	ESE	10.0	Prec_Ext
05/01/2017	28.6	72.0	10.8	813	SE	9.0	Prec
06/01/2017	24.9	74.0	1.2	815	SE	5.0	NA
07/01/2017	30.4	69.0	4.9	812	SE	7.0	Prec_Ext
08/01/2017	31.0	66.0	0.0	810	ESE	7.0	Prec_Ext
09/01/2017	28.7	69.0	0.0	809	WSW	4.0	Prec
10/01/2017	26.2	87.0	16.9	814	ESE	9.0	NA
11/01/2017	26.7	69.0	0.0	815	ENE	2.0	Prec
12/01/2017	28.4	61.0	0.0	816	E	5.0	Prec
13/01/2017	32.6	53.0	0.0	813	SE	9.0	Prec_Ext
14/01/2017	32.0	52.0	0.0	812	SE	9.0	Prec_Ext
15/01/2017	32.6	54.0	25.2	809	E	7.0	Prec_Ext

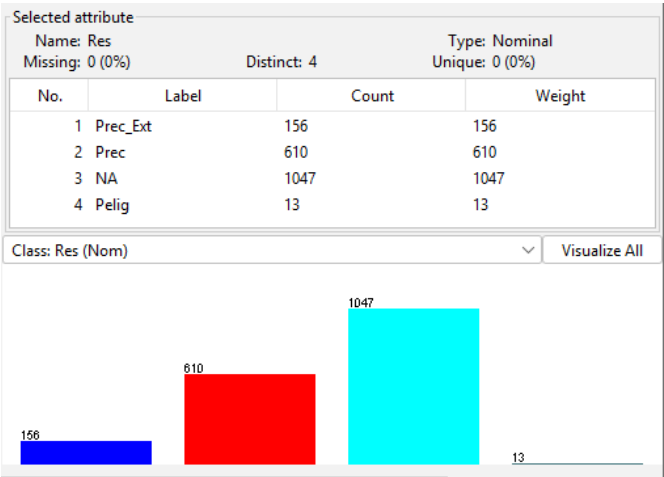
Fuente: Elaboración propia

La Tabla 3 cuenta con 1826 instancias y 8 atributos correspondiente al registro diario de variables meteorológicas.

La Figura 11 muestra la distribución de las instancias agrupadas por el atributo Res (Nivel de riesgo): Prec_Ext, Prec, NA y Pelig con 156, 610, 1047 y 13 instancias respectivamente.

Asimismo, se establecen también los nombres de los atributos de entrada para su procesamiento en la herramienta Weka: temperatura máxima (Tm), humedad relativa (Hr), precipitación (Pp), presión atmosférica (Pat), dirección del viento (Dv) y velocidad del viento (Vv); además, se incluye el atributo Mes con el fin de evaluar si el mismo incide en el árbol de decisión resultante.

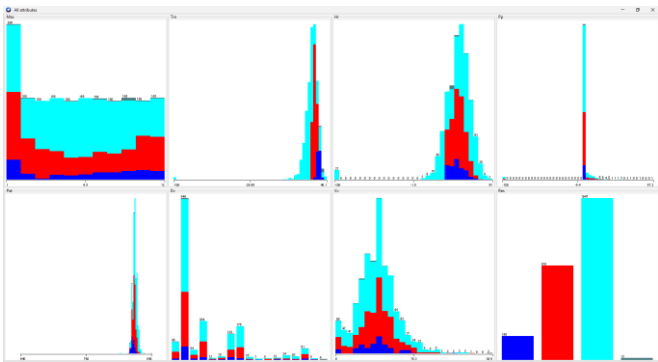
Figura 11. Distribución de instancias agrupadas por el nivel de riesgo (Res)



Fuente: Elaboración propia con la herramienta Weka

La Figura 12 muestra la distribución de las instancias de cada uno de los atributos de entrada identificados.

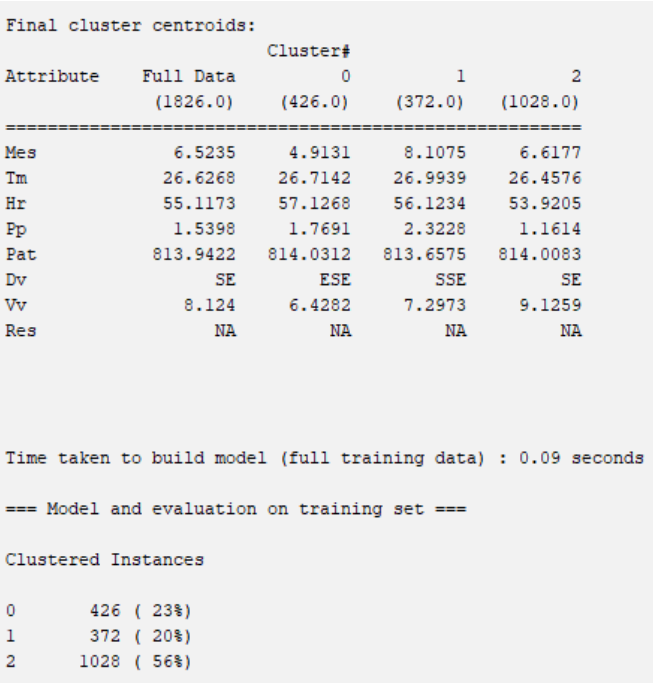
Figura 12. Distribución de instancias por atributos



Fuente: Elaboración propia con la herramienta Weka

La Figura 13 muestra la distribución de las instancias en tres clusters, es decir, la manera en la cual fueron clasificados todos los atributos de entrada (variables meteorológicas), donde el cluster 0, es el grupo que presenta un 23%; el cluster 1 presenta un 20% y el cluster 2 presenta un 56%.

Figura 13. Distribución de instancias por clusters

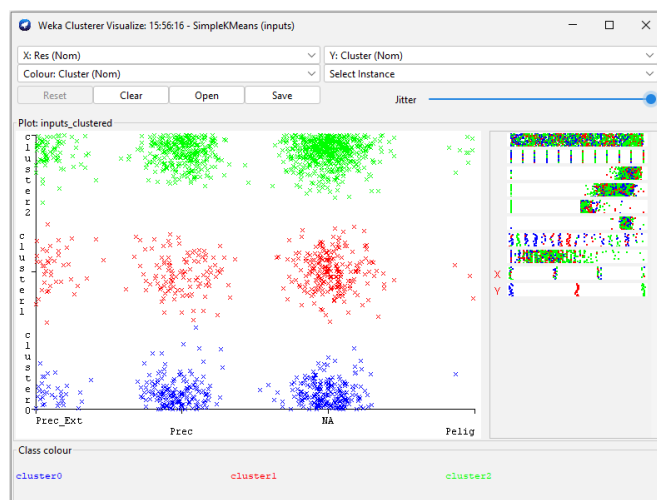


Fuente: Elaboración propia con la herramienta Weka

Por otra parte, la Figura 14 muestra el diagrama de dispersión del atributo de salida o resultado (nivel de riesgo) en cada uno de los clusters aplicando el algoritmo K-medias, donde se observa una mayor concentración del resultado NA (No aplica), en tamaño medio corresponde al nivel de riesgo de precaución, el siguiente tamaño es de precaución extrema y finalmente de peligro con el conjunto agrupado con pocas instancias. Estos resultados nos dan una primera pauta de cómo estará distribuido porcentualmente el

resultado final de nivel de riesgo por estrés térmico por calor.

Figura 14. Distribución del atributo de salida de nivel de riesgo (Res) en cada uno de los clusters



Fuente: Elaboración propia con la herramienta Weka

Asimismo, la Figura 15 y Figura 16 muestran la distribución de los atributos que corresponden a la temperatura máxima y la humedad relativa por cada uno de los clusters generados. En el caso de la temperatura máxima, la mayor concentración se presenta con los valores superiores 21 °C y menores a 35° C y en el caso de la humedad relativa se muestra una mayor concentración entre los valores de 35% y 75%.

Figura 15. Distribución del atributo de temperatura máxima (Tm) en cada uno de los clusters



Fuente: Elaboración propia con la herramienta Weka

Figura 16. Distribución del atributo de humedad relativa (Hr) en cada uno de los clusters



Fuente: Elaboración propia con la herramienta Weka

Después de aplicar el algoritmo de clustering, se utilizó el algoritmo C4.5 (J48) cuyo resultado se muestra a través de un resumen, donde se especifican las instancias clasificadas y no clasificadas correctamente, además de una matriz de confusión que indica para cada clase la cantidad de instancias clasificadas por el atributo de resultado.

En la Figura 17, se muestra que se tienen 1812 instancias clasificadas correctamente, lo cual representa un 99.233% de aciertos y 14 instancias clasificadas incorrectamente, lo cual representa un 0.77% de desaciertos; teniendo así un buen porcentaje de aciertos y alto grado de confiabilidad.

Figura 17. Matriz de confusión

```

=== Summary ===
Correctly Classified Instances      1812      99.2333 %
Incorrectly Classified Instances    14         0.7667 %
Kappa statistic                    0.9861
Mean absolute error                 0.006
Root mean squared error             0.0555
Relative absolute error              2.1574 %
Root relative squared error         14.9404 %
Total Number of Instances          1826

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,962	0,003	0,968	0,962	0,965	0,961	0,996	0,989	Prec_Ext
	0,998	0,003	0,993	0,998	0,996	0,994	0,999	0,997	Prec
	0,995	0,000	1,000	0,995	0,998	0,994	0,999	1,000	NA
	0,846	0,003	0,688	0,846	0,759	0,761	0,998	0,863	Pelig
Weighted Avg.	0,992	0,001	0,993	0,992	0,993	0,990	0,999	0,997	

```

=== Confusion Matrix ===
a b c d <-- classified as
150 1 0 5 | a = Prec_Ext
1 609 0 0 | b = Prec
2 3 1042 0 | c = NA
2 0 0 11 | d = Pelig

```

Fuente: Elaboración propia con la herramienta Weka

Posteriormente, en la Figura 18 se visualiza el árbol de decisión generado por la herramienta Weka.

A continuación, se tiene el detalle de las reglas de clasificación generadas en estrecha relación con el

árbol de decisión resultante. Cabe señalar que el orden en el cual se presentan las reglas es mutuamente excluyente y exhaustivo.

Regla 0:

Tm <= 26.6: NA

Regla 1:

Tm > 26.6
| Hr <= 39.3415: NA

Regla 2:

Tm > 26.6
| Hr > 39.3415
| | Tm <= 31.1
| | | Tm <= 29.9: Prec

Regla 3:

Tm > 26.6
| Hr > 39.3415
| | Tm <= 31.1
| | | Tm > 29.9
| | | | Hr <= 59: Prec

Regla 4:

Tm > 26.6
| Hr > 39.3415
| | Tm <= 31.1
| | | Tm > 29.9
| | | | Hr > 59: Prec_Ext

Regla 5:

Tm > 26.6
| Hr > 39.3415
| | Tm > 31.1
| | | Tm <= 32.2
| | | | Hr <= 49.3557: Prec

Regla 6:

Tm > 26.6
| Hr > 39.3415
| | Tm > 31.1
| | | Tm <= 32.2

| | | | Hr > 49.3557: Prec_Ext

Regla 7:

Tm > 26.6
| Hr > 39.3415
| | Tm > 31.1
| | | Tm > 32.2
| | | | Tm <= 35.5
| | | | | Hr <= 54.3628: Prec_Ext

Regla 8:

Tm > 26.6
| Hr > 39.3415
| | Tm > 31.1
| | | Tm > 32.2
| | | | Tm <= 35.5
| | | | | Hr > 54.3628
| | | | | Tm <= 33.6: Prec_Ext

Regla 9:

Tm > 26.6
| Hr > 39.3415
| | Tm > 31.1
| | | Tm > 32.2
| | | | Tm <= 35.5
| | | | | Hr > 54.3628
| | | | | Tm > 33.6: Pelig

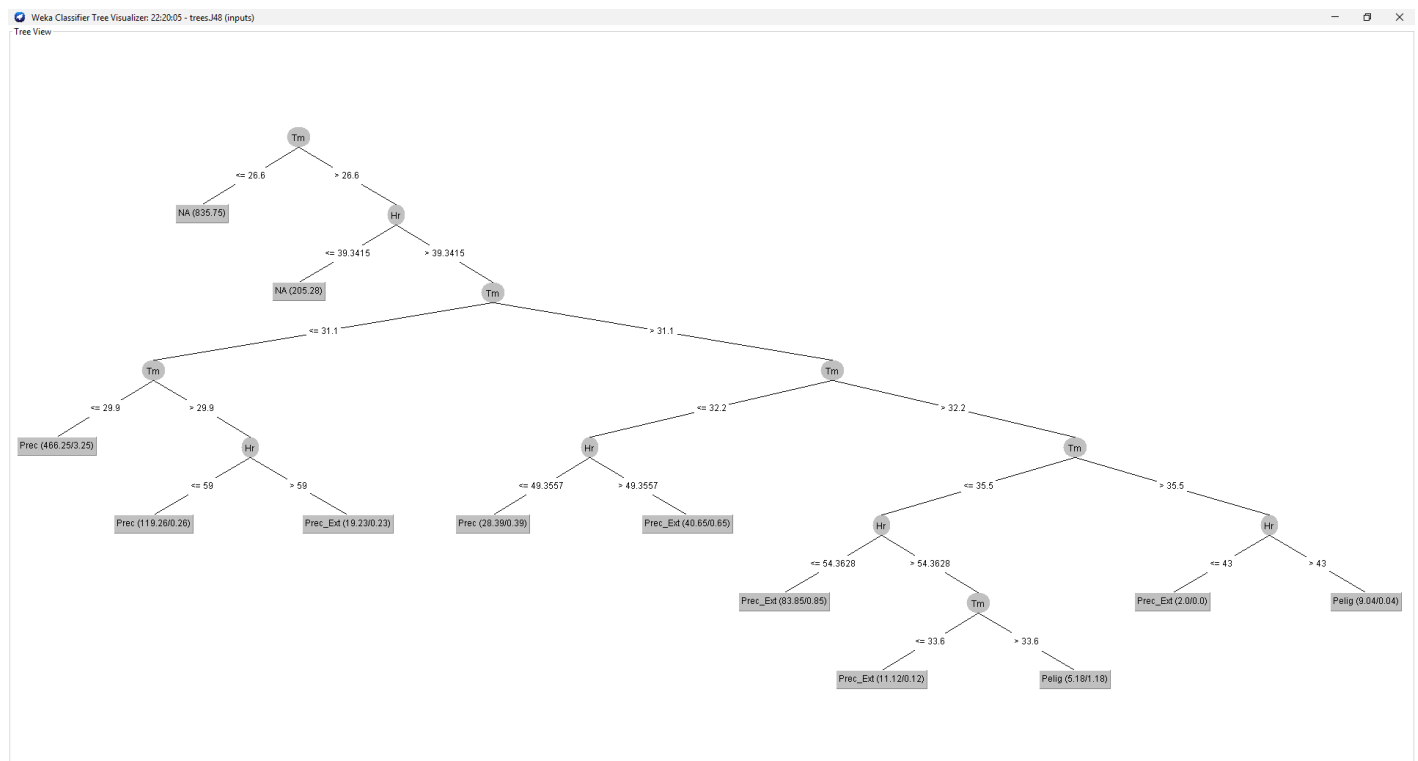
Regla 10:

Tm > 26.6
| Hr > 39.3415
| | Tm > 31.1
| | | Tm > 32.2
| | | | Tm > 35.5
| | | | | Hr <= 43: Prec_Ext

Regla 11:

Tm > 26.6
| Hr > 39.3415
| | Tm > 31.1
| | | Tm > 32.2
| | | | Tm > 35.5
| | | | | Hr > 43: Pelig

Figura 18. Árbol de decisión resultante



Fuente: Elaboración propia con la herramienta Weka

DISCUSIÓN

Para la generación del árbol de decisión y sus correspondientes reglas de clasificación, si bien se utilizaron las variables meteorológicas de temperatura máxima, humedad relativa, precipitación, presión atmosférica, dirección del viento y velocidad del viento, además del mes de cada instancia, se pudo observar que solo dos variables son las que inciden en la determinación del nivel de riesgo del estrés térmico por calor, tal como plantea el Servicio Meteorológico Nacional (MWS): temperatura (máxima) y humedad relativa.

Con respecto a la inclusión de las demás variables para la generación de las reglas de clasificación, se pretendía determinar si es que estas incidían de alguna manera en el resultado; sin embargo, por los resultados obtenidos se confirma que no son necesarias para el caso específico de estudio en los años 2017 y 2022, pudiendo definir los niveles de riesgo con dos variables.

Por otra parte, también se pudo establecer que la ciudad de Tarija no presenta resultados de peligro extremo, por lo que no es necesario aplicar medidas de protección aún más potentes. Sin embargo, se tienen 2 reglas con nivel de riesgo de peligro (0.61% de instancias clasificadas correctamente), lo que implica tomar medidas de precaución para los trabajadores que realizan actividades al aire libre, aunque su nivel de ocurrencia es bastante bajo; 5 reglas con nivel de riesgo de precaución extrema (8.28% de instancias clasificadas correctamente), lo que significa que se debe implementar precauciones y aumentar la conciencia de cuidado; 3 reglas con nivel de riesgo de precaución (33.61% de instancias clasificadas correctamente), las cuales pueden controlarse a través de la planificación y seguridad básica contra el calor.

Por otra parte, se tienen 2 reglas NA (No aplica), lo cual significa que –por los valores de entrada (temperatura máxima y humedad relativa)– no implica riesgo para los trabajadores, teniendo un alto porcentaje de instancias clasificadas correctamente en este nivel con un 57.50%.

La Organización Internacional del Trabajo (OIT, 2019) a través de su publicación “Trabajar en un Planeta más Caliente. El Impacto del Estrés Térmico en la Productividad Laboral y el Trabajo Decente”, utiliza una metodología que combina modelos climáticos, datos meteorológicos con previsiones de fuerza laboral y datos de salud en el trabajo con el fin de determinar el impacto del estrés térmico en la actividad laboral.

Este estudio indica que la productividad laboral en América del Sur se ve afectada por el estrés térmico y por ende en el Estado Plurinacional de Bolivia, donde realiza una relación del total de puestos de trabajos a

tiempo completo en función a las horas de trabajo perdidas por el estrés térmico de los años 1995 y 2030, donde esta cifra sube de 3.5 a 25 mil respectivamente. La metodología de la OIT utiliza uno de los índices de estrés térmico más comunes en salud profesional: la temperatura de bulbo húmedo y de globo (WBGT), cuya medida se calcula sobre la base de la temperatura, la humedad, la velocidad del viento y el calor irradiado del mes más caluroso en áreas geográficas pequeñas. Asimismo, relaciona los peligros para la salud laboral asociados al calor según el Quinto Informe de Evaluación del Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC).

Este estudio concluye que los niveles de consumo metabólico o intensidad laboral deben evitarse cuando la temperatura corporal sea mayor a 38 °C, donde la aclimatación fisiológica al calor es una medida de protección, aunque no aplicable de manera inmediata, sino a partir de una a dos semanas de aclimatación al calor.

Ahora, con relación al presente trabajo, en el estudio de la OIT se establece un umbral para activar los mecanismos de precaución laboral por estrés térmico por calor, el cual tiene que ver con la temperatura corporal según el área geográfica donde se encuentren los trabajadores; en cambio, en la investigación realizada se establecen cuatro niveles para la identificación del nivel de riesgo asociado, donde se generan reglas que relacionan la temperatura máxima y la humedad relativa para establecer a qué nivel corresponde y cuál es la medida de protección recomendada. Adicionalmente, el estudio de la OIT incorpora el análisis de las horas de trabajo perdidas por el estrés térmico por sector y por país.

Por otra parte, de acuerdo a la investigación realizada por Aquino, M. et al. (2019), cuyo trabajo lleva por título “Determinación de los Intervalos de Confort y Estrés Térmico para Espacios Abiertos en la Ciudad de Maringá (PR), Brasil, utilizando el Índice PET”, analiza la sensación térmica real de la ciudad de Maringá a partir del índice de confort térmico PET en función a variables climáticas, personales y subjetivas, cuyo resultado inicial presenta un bajo porcentaje de aciertos (67%), procediendo a calibrar la misma a través de votos de sensación real de las personas, teniendo como resultado final la propuesta de intervalos de clasificación de estrés térmico a partir de la temperatura del aire (T_a), donde:

- Estrés térmico para frío, $T_a \leq 19.5^\circ\text{C}$
- Confort, $19.5 < T_a < 30^\circ\text{C}$
- Estrés térmico para calor, $T_a \geq 30^\circ\text{C}$

Realizando la correspondiente comparativa con la investigación, en el presente trabajo se realiza el

análisis de estrés térmico por calor sobre las instancias identificadas en la ciudad de Tarija basado en la herramienta de índice de calor del NWS y la técnica de minería de datos de árboles de decisión para la ciudad de Tarija, donde se considera la temperatura del aire mayor o igual a 26.7°C y la humedad relativa mayor o igual a 40%, cuya regla es que mientras más altos sean estos valores, más alto es el nivel de riesgo y por ende las medidas de protección son más potentes.

Asimismo, se establecen reglas de clasificación para poder determinar cuándo se presenta un cierto nivel de riesgo en función a los valores de entrada de la temperatura máxima del aire y la humedad relativa. Finalmente, en el presente trabajo, solo se consideran las variables meteorológicas como insumos de análisis.

REFERENCIAS BIBLIOGRÁFICAS

3M. Ciencia aplicada a la vida. (2017). *Protección y Seguridad Industrial*. Fecha de consulta: 1 de octubre de 2022, de https://www.3m.com.bo/3M/es_BO/epp-la/soporte-EPP/tips-seguridad-industrial/proteccion-estres-termico/

Aguirre, H. y Rincón, N. (2015). Minería de Procesos: Desarrollo, Aplicaciones y Factores Críticos, *Cuadernos de Administración*. [En línea]. Vol. 28, N° 50, noviembre 2015, pp 137 – 157. Fecha de consulta: 1 de septiembre de 2022. Disponible en: https://www.researchgate.net/profile/Santiago-Aguirre-2/publication/285383521_Mineria_de_procesos_desarrollo_aplicaciones_y_factores_criticos/links/56afb25b08ae9c1968b46ca2/Mineria-de-procesos-desarrollo-aplicaciones-y-factores-criticos.pdf

Aquino, M. et al. (2019). Determinación de los Intervalos de Confort y Estrés Térmico para Espacios Abiertos en la Ciudad de Maringá (PF), Brasil, utilizando el Índice PET, *Desenvolvimento e meio ambiente*. [En línea]. Vol. 51, Agosto de 2019, pp 311 – 327. Fecha de consulta: 1 de octubre de 2022. Disponible en: <https://revistas.ufpr.br/made/article/view/60679/39385>

Barcía, S. et al. (2021). Los Extremos Climáticos por Calor, *Revista Cubana de Meteorología*. [En línea]. Vol. 27, N° 1, Enero – Marzo 2021, pp 1 – 16. Fecha de consulta: 21 de septiembre de 2022. Disponible en: https://www.researchgate.net/profile/Sinai-Barcia-Sardinas/publication/354776114_Los_extremos_climaticos_por_calor/links/614bf6e5519a1a381f795cf7/Los-extremos-climaticos-por-calor.pdf

Bouza, C. y Santiago, A. (2012). La Minería de Datos: Árboles de Decisión y su Aplicación en Estudios Médicos, *Modelación Matemática de Fenómenos del Medio Ambiente y la Salud*. [En línea]. Tomo 2, pp 64 – 78. Fecha de consulta: 24 de septiembre de 2022.

Disponible en: https://rideca.cs.buap.mx/web/files/articulo_itBUoUwI_AaJENf.pdf

Caballero, L., et al. (2015). Árbol de Decisión C4.5 basado en Entropía Minoritaria para Clasificación de Conjuntos de Datos no Balanceados, *Research in Computing Science*. [En línea]. Vol. 92, mayo 2015, pp 23-34. Fecha de consulta: 11 de septiembre de 2022. Disponible en: https://www.rcs.cic.ipn.mx/2015_92/Arbol%20de%20decision%20C4_5%20basado%20en%20entropia%20minoritaria%20para%20clasificacion%20de%20conjuntos%20de%20datos.pdf

El País. (2019). *Cambio Climático Costará 80 Millones de Puestos de Trabajo en 2030*. [Periódico en línea]. Fecha de consulta: 2 de octubre de 2022, de https://elpais.bo/internacional/20190719_cambio-climatico-costara-80-millones-de-puestos-de-trabajo-en-2030.html

ESRI. (s.f.). *Qué es ArcMap*. [En línea]. Fecha de consulta: 21 de septiembre de 2022, de <https://desktop.arcgis.com/es/arcmap/10.3/main/map/what-is-arcmap-.htm>

GeoBolivia. (s.f.). *Infraestructura de Datos Espaciales del Estado Plurinacional de Bolivia* [Repositorio en línea]. Fecha de consulta: 16 de septiembre de 2022. Disponible en: <http://geo.gob.bo/porta/>

Gobierno Autónomo Departamental de Tarija (GADT). (2014). Plan Departamental de Gestión del Riesgo y Adaptación al Cambio Climático (2014-2018). [Plan en línea]. Fecha de consulta: 2 de octubre de 2022. Disponible en: <http://201.222.81.2/AnalisisNormativo/normasPDF/5934.pdf>

Instituto de Hidráulica Ambiental de la Universidad de Cantabria (IHCantabria). (2021). *Índice de Vulnerabilidad y Adaptación al Cambio Climático en la Ciudad de Tarija, Bolivia. Resumen Ejecutivo*. [Libro en línea]. Fecha de consulta: 29 de septiembre de 2022. Disponible en: <http://scioteca.caf.com/handle/123456789/1815>

Linares, C. et al. (2017). *Temperaturas Extremas y Salud. Cómo nos Afectan las Olas de Calor y Frío*. [Libro en línea]. Instituto de Salud Carlos III. Fecha de consulta: 18 de septiembre de 2022. Disponible en: <http://gesdoc.isciii.es/gesdoccontroller?action=download&id=11/02/2020-aae7e5ae3a>

Ministerio de Trabajo y Asuntos Sociales España (MTASE). (s.f.). *Prevención de Riesgos Laborales debido al Estrés Térmico por Calor*. [En línea]. Fecha de consulta: 26 de septiembre de 2022. Disponible en:

<https://www.navarra.es/NR/rdonlyres/AF2BD786-0A6D-4564-9076-BE42220B4843/225685/calorytrabajoprofesional.pdf>

Ministerio de Trabajo, Migraciones y Seguridad Social – España (MTMSS). (2018). *Golpe de Calor en Trabajos al Aire Libre*. [Libro en línea]. CCOO Aragón. Fundación Estatal para la Prevención de Riesgos Laborales AR 2018-0043. Fecha de consulta: 16 de septiembre de 2022. Disponible en: https://www.saludlaboralymedioambiente.ccooaragon.com/documentacion/gs2019_Golpe-calor%20ccoo%20web.pdf

National Weather Service (NWS). (s.f.a). *Heat Forecast Tools*. Fecha de consulta: 20 de septiembre de 2022, de <https://www.weather.gov/safety/heat-index>

National Weather Service (NWS). (s.f.b). *What is the Heat Index?*. Fecha de consulta: 21 de septiembre de 2022, de <https://www.weather.gov/safety/heat-index>

Organización Internacional del Trabajo (OIT). (2019). *Trabajar en un Planeta más Caliente: El Impacto del Estrés Térmico en la Productividad Laboral y el Trabajo Decente*. [Libro en línea]. Fecha de consulta: 13 de octubre de 2022. Disponible en: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_768707.pdf

Pérez, J. et al. (2018). Una Heurística Eficiente Aplicada al Algoritmo K-means para el Agrupamiento de Grandes Instancias Altamente Agrupadas, *Computación y Sistemas*. [En línea]. Vol. 22, N° 2, pp 607 –619. Fecha de consulta: 14 de septiembre de 2022. Disponible en: <https://www.cys.cic.ipn.mx/ojs/index.php/CyS/article/view/2548/2476>

Salazar, A. y Álvarez, L. (2020). El Índice de Calor: un Factor de Alerta Temprana en Salud Pública y Ciudades Sostenibles, *Revista Salud Uninorte*. [En línea]. Vol. 35, N° 3, pp 440 – 449. Fecha de consulta: 27 de septiembre de 2022. Disponible en: <https://rcientificas.uninorte.edu.co/index.php/salud/articloe/view/12278/214421445019>

Servicio Nacional de Meteorología e Hidrología (SENAMHI). (s.f.). *Información Nacional de Datos Hidrometeorológicos*. [Sistema en línea]. Fecha de consulta: 20 de septiembre de 2022. Disponible en: <http://senamhi.gob.bo/index.php/sysparametros>

Valcárcel, V. (2004). Data Mining y el Descubrimiento del Conocimiento, *Industrial Data*. [En línea]. Vol. 7, N° 2, pp 83 – 86. Fecha de consulta: 28 de agosto de 2022. Disponible en: <https://www.redalyc.org/articulo.oa?id=81670213>

Artículo

Recibido: 14 de octubre de 2022

Aceptado: 25 de noviembre de 2022

Cita sugerida:

Vásquez, N.S. (2022). Reglas de clasificación de estrés térmico por calor a partir de árboles de decisión para la ciudad de Tarija. *Revista SEC CIENCIA*. 4(6), 8-19. <http://repo.uajms.edu.bo/index.php/secciencia/issue/view/4>