



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE CIENCIAS

CARRERA ESTADÍSTICA

**PREDICCIÓN DE PATRONES DE COMPORTAMIENTO DE LA
TEMPERATURA MEDIANTE HERRAMIENTAS ESTADÍSTICAS
USADAS EN LOS MODELOS DE CIRCULACIÓN ATMOSFÉRICA
2020-2050**

Trabajo de Titulación

Tipo: Proyecto de Investigación

Presentado para obtener al grado académico de:

INGENIERA EN ESTADÍSTICA INFORMÁTICA

AUTORA:

ERIKA ALEXANDRA TOAPANTA YUGCHA

Riobamba – Ecuador

2022



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

**PREDICCIÓN DE PATRONES DE COMPORTAMIENTO DE LA
TEMPERATURA MEDIANTE HERRAMIENTAS ESTADÍSTICAS
USADAS EN LOS MODELOS DE CIRCULACIÓN ATMOSFÉRICA
2020-2050**

Trabajo de Titulación

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

INGENIERA EN ESTADÍSTICA INFORMÁTICA

AUTORA: ERIKA ALEXANDRA TOAPANTA YUGCHA

DIRECTORA: Ing. CRISTINA ESTEFANIA RAMOS ARAUJO MSc.

Riobamba – Ecuador

2022

©2022, Erika Alexandra Toapanta Yugcha

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.

Yo, ERIKA ALEXANDRA TOAPANTA YUGCHA, declaro que el presente Trabajo de Titulación es de mi autoría y los resultados de este son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autor asumo la responsabilidad legal y académica de los contenidos de este Trabajo de Titulación; el patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 24 noviembre de 2022



Erika Alexandra Topanta Yugcha

070442762-4

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

El Tribunal del Trabajo de Titulación certifica que: El Trabajo de Titulación; Tipo: Proyecto de Investigación, **PREDICCIÓN DE PATRONES DE COMPORTAMIENTO DE LA TEMPERATURA MEDIANTE HERRAMIENTAS ESTADÍSTICAS USADAS EN LOS MODELOS DE CIRCULACIÓN ATMOSFÉRICA 2020-2050**, realizado por la señorita: **ERIKA ALEXANDRA TOAPANTA YUGCHA**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Titulación, el mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal Autoriza su presentación.

	FIRMA	FECHA
Ing. Natalia Alexandra Perez Londo MSc. PRESIDENTE DEL TRIBUNAL	 _____	2022-11-24
Ing. Cristina Estefanía Ramos Araujo MSc. DIRECTORA DEL TRABAJO DE TITULACIÓN	 _____	2022-11-24
Ing. Johanna Enith Aguilar Reyes.MSc. ASESORA DEL TRABAJO DE TITULACIÓN	 _____	2022-11-24

DEDICATORIA

El trabajo realizado es dedicado a mi familia en especial a mis padres, por el apoyo que me brindaron durante mi carrera estudiantil, apoyándome en los momentos más difíciles de mi vida y de mi carrera les agradezco por cada lección que fue un aprendizaje para poder lograr mis objetivos y mis metas planeadas que hoy es una realidad, a mis hermanos Jorge, Vanessa, Kevin, Emylis , que han sido mi motivación para nunca rendirme y seguir firme en cada paso dado.

Erika

AGRADECIMIENTO

Un agradecimiento especial a la Ing. Cristina Ramos, directora de mi tesis y a la Ing. Johanna Aguilar, quienes me han brindado su apoyo incondicional con sus conocimientos, tiempo y paciencia lo cual me permitió culminar con éxito mi tesis, así también por compartir sus conocimientos como docentes de algunas materias a lo largo de mi vida universitaria.

A todos los docentes de la carrera de Estadística que semestre a semestre fueron impartiendo valiosos conocimientos que me han ayudado a formarme para llegar a ser una excelente profesional.

Erika

TABLA DE CONTENIDO

ÍNDICE DE TABLAS.....	ix
ÍNDICE DE FIGURAS.....	x
ÍNDICE DE GRÁFICOS.....	xi
ÍNDICE DE ANEXOS	¡Error! Marcador no definido.
RESUMEN.....	¡Error! Marcador no definido.
ABSTRACT.....	¡E rror! Marcador no definido.
INTRODUCCIÓN	1

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL.....	7
1.1. Bases conceptuales.....	7
<i>1.1.1. Estadística</i>	<i>7</i>
<i>1.1.2. Medidas tendencia central.....</i>	<i>7</i>
<i>1.1.3. Desviación estándar.....</i>	<i>8</i>
<i>1.1.4. Varianza</i>	<i>8</i>
<i>1.1.5. Gráficas Estadísticas</i>	<i>8</i>
<i>1.1.6. Aprendizaje no supervisado.....</i>	<i>9</i>
<i>1.1.7. Reducción de dimensionalidad.....</i>	<i>9</i>
<i>1.1.8. Agrupamiento o clustering.....</i>	<i>11</i>
<i>1.1.9. K-means.....</i>	<i>12</i>
<i>1.1.10. Distancias.....</i>	<i>12</i>
<i>1.1.11. Clustering Jerárquico.....</i>	<i>12</i>
<i>1.1.12. Índice Silhouette.....</i>	<i>13</i>
<i>1.1.13. Matriz de confusión.....</i>	<i>13</i>
<i>1.1.14. Almacenamiento de datos NetCDF.....</i>	<i>13</i>
<i>1.1.15. Dataset.....</i>	<i>14</i>
<i>1.1.16. Anaconda.....</i>	<i>15</i>
<i>1.1.17. Júpiter.....</i>	<i>15</i>
1.2. Bases teóricas	15
<i>1.2.1. Cambio climático</i>	<i>15</i>
<i>1.2.2. Meteorología.....</i>	<i>15</i>
<i>1.2.3. Componentes del sistema climático.....</i>	<i>16</i>
<i>1.2.4. Topografía y geología.....</i>	<i>16</i>

1.2.5.	<i>Clima</i>	17
1.2.6.	<i>Microclima</i>	17
1.2.7.	<i>Eventos extremos</i>	17
1.2.8.	<i>Circulación general de la atmosfera</i>	18
1.2.9.	<i>Temperatura</i>	18
1.2.10.	<i>Variabilidad natural</i>	18
1.2.11.	<i>Radiación terrestre</i>	19
1.2.12.	<i>Efecto invernadero</i>	19
1.2.13.	<i>Variaciones de temperatura</i>	19
1.2.14.	<i>Anomalía climática</i>	19
1.2.15.	<i>Temperatura media global</i>	20

CAPÍTULO II

2.	MARCO METODOLÓGICO	21
2.1.	Tipo de investigación	21
2.2.	Diseño de la Investigación	23
2.2.1.	<i>Localización del estudio</i>	23
2.2.2.	<i>Población de estudio</i>	24
2.2.3.	<i>Tamaño de la muestra</i>	24
2.2.4.	<i>Método de muestreo</i>	24
2.2.5.	<i>Técnica de recolección de los datos</i>	24
2.2.6.	<i>Modelo estadístico</i>	24

CAPÍTULO III

3.	MARCO DE RESULTADOS Y DISCUSIÓN DE LOS RESULTADOS	25
2.1.	Descripción de la matriz de información	25
3.1.	Estadísticas descriptivas	26
3.1.1.	<i>Exploración de la variable temperatura</i>	27
3.1.2.	<i>Matriz bidimensional</i>	27
3.2.	Reducción de la dimensionalidad	28
3.2.1.	<i>Componentes principales</i>	28
3.2.2.	<i>Diagrama de dispersión</i>	29
3.2.3.	<i>T-SNE</i>	30
3.2.4.	<i>UMAP</i>	32
3.3.1.	<i>Algoritmo K Means - UMAP con parámetros (métrica="chebyshev")</i>	34

3.3.2.	<i>Agrupación de datos según el rango de años</i>	36
3.3.3.	<i>Matriz de confusión</i>	36
3.3.4.	<i>Representación de resultados por días del algoritmo K-means métrica chebyshev.</i>	37
3.4.	Mapas obtenidos	38
3.4.1.	<i>Temperaturas Altas (2031-2035)</i>	41
3.4.2.	<i>Temperaturas Bajas (2036-2045)</i>	43
3.5.	Algoritmos que fueron probados para su selección	46
3.5.1.	<i>Algoritmo K- means - UMAP con parámetros (métrica="euclidean")</i>	46
3.5.2.	<i>Algoritmo DBSCAN - UMAP con parámetros (metric="euclidean")</i>	46
3.5.3.	<i>Algoritmo DBSCAN - UMAP con parámetros (métrica="chebyshev")</i>	47
3.5.4.	<i>Algoritmo AGGLOMERATIVE CLUSTERING - UMAP con parámetros (métrica="euclidean")</i>	48
3.5.5.	<i>Algoritmo AGGLOMERATIVE CLUSTERING - UMAP con parámetros (métrica="chebyshev")</i>	49
3.5.6.	<i>Validación del modelo</i>	49
3.5.7.	<i>Validacion del modelo</i>	51
CONCLUSIONES		53
RECOMENDACIONES		54
BIBLIOGRAFÍA		
ANEXOS		

ÍNDICE DE TABLAS

Tabla 1-2:	Operacionalización de variables	24
Tabla 1-3:	Tabla de la variable Temperatura bajo sus tres dimensiones	25
Tabla 2-3:	Estadísticos calculados por promedio de filas.	26
Tabla 3-3:	Dimensiones de la matriz bidimensional	27
Tabla 4-3:	Rango de años para los grupos.	36
Tabla 5-3:	Matriz de confusión	36
Tabla 6-3:	Rango de años para los grupos.	38
Tabla 7-3:	Índices de Silhouette.....	50
Tabla 8-3:	Temperaturas registradas en el Ecuador mes de Mayo del año 2021.	51
Tabla 9-3:	Temperaturas registradas en el Ecuador mes de Junio del año 2021.....	51
Tabla 10-3:	Temperaturas registradas en el Ecuador mes de Julio del año 2021.....	51
Tabla 11-3:	Temperaturas registradas en el Ecuador mes de Agosto del año 2021.....	52

ÍNDICE DE FIGURAS

Figura 1-1:	Matriz de confusión para una clasificación binaria.....	11
Figura 2-1:	Datos de un área que varían con el tiempo	14
Figura 3-1:	Estructura de un Dataset del paquete xarray	14
Figura 4-1:	Componentes del Sistema climático	16
Figura 1-2:	Proceso cuantitativo	232
Figura 2-2:	Mapa del Ecuador	23

ÍNDICE DE GRÁFICOS

Gráfico 1-1:	Componentes principales.....	10
Gráfico 2-1:	Algoritmo TSNE.....	10
Gráfico 3-1:	Algoritmo UMAP.....	11
Gráfico 4-1:	Agrupamiento de datos en 3 grupos diferentes.....	12
Gráfico 1-3:	Histograma de frecuencias de las anomalías.	26
Gráfico 2-3:	Gráfico de líneas del promedio de filas por año.	27
Gráfico 3-3:	Elección del número de componentes principales.....	28
Gráfico 4-3:	Porcentaje de varianza explicada.....	29
Gráfico 5-3:	Diagrama de dispersión de las dos primeras componentes.....	29
Gráfico 6-3:	Diagrama de dispersión de las dos primeras componentes por mes.....	30
Gráfico 7-3:	Incrustación de vecinos estocásticos distribuidos en t	31
Gráfico 8-3:	Incrustación de vecinos respecto a los años.	31
Gráfico 9-3:	Umap dimensionalidad	32
Gráfico 10-3:	Umap clasificado por años	33
Gráfico 11-3:	Gráfica de codo.....	34
Gráfico 12-3:	Umap métrica chebyshev respecto meses.....	34
Gráfico 13-3:	Umap métrica chebyshev respecto años.	35
Gráfico 14-3:	Umap métrica chebyshev agrupación K-means	35
Gráfico 15-3:	Diagrama de dispersión UMAP.....	39
Gráfico 16-3:	Diagrama de frecuencias temperaturas.....	38
Gráfico 17-3:	Mapas generados con la temperatura promedio para cada año.	39
Gráfico 18-3:	Gráfico de Calor representado diariamente respecto a los años	40
Gráfico 19-3:	Mapa generado con la temperatura promedio correspondiente al 2031.	41
Gráfico 20-3:	Mapa generado con la temperatura promedio correspondiente al 2032.....	41
Gráfico 21-3:	Mapa generado con la temperatura promedio correspondiente al 2034.....	42
Gráfico 22-3:	Grafico de Calor representado en los últimos 4 años	42
Gráfico 23-3:	Mapa generado con la temperatura promedio correspondiente al 2036.....	43
Gráfico 24-3:	Mapa generado con la temperatura promedio correspondiente al 2039.....	44
Gráfico 25-3:	Mapa generado con la temperatura promedio correspondiente al 2038.....	44
Gráfico 26-3:	Grafico de Calor representado en los últimos 4 años	45
Gráfico 27-3:	Umap métrica euclídea respecto años.....	46
Gráfico 28-3:	Umap métrica euclídea agrupación K-means	46
Gráfico 29-3:	Umap métrica euclídea agrupación DBSCAN.	49
Gráfico 30-3:	Umap métrica euclídea agrupación DBSCAN.	49

Gráfico 31-3: Umap métrica euclídea agrupación Agglomerative Clustering.....	48
Gráfico 32-3: Umap métrica euclídea agrupación Agglomerative Clustering.....	49
Gráfico 33-3: Diagrama de barras representando a los algoritmos y métricas.....	50

ÍNDICE DE ANEXOS

ANEXO A: OFICIO DE USO DE LA INFORMACIÓN DEL INHAMI

RESUMEN

El trabajo de Titulación tuvo como objetivo predecir los patrones de comportamiento de la temperatura, donde se tomó como población de estudio todas las provincias del Ecuador, existen pocos análisis dedicados a predecir patrones de comportamiento, es por esto que se estudió la información que se obtuvo de los modelos de circulación atmosférica a futuro del banco de datos CHIRPS, esta información al estar en forma de matriz multidimensional se tuvo que transformar para poder trabajar con ella como matriz bidimensional, al realizar un análisis exploratorio se encontró una tendencia creciente no muy pronunciada pero clara, donde se evidencia que con el transcurso del tiempo la temperatura tendrá un incremento, además sus datos provienen de una distribución normal, luego de aplicar varios algoritmos de reducción de dimensionalidad. Como resultado se encontró que el algoritmo UMAP es mejor para encontrar una representación en dos dimensiones de esta matriz con varias columnas en la presente investigación, ayudando a conservar su estructura topológica multidimensional, así mismo se utilizó la distancia de Chebyshev para obtener mejores resultados, en donde el algoritmo efectuado para realizar las agrupaciones k-means resultó el mejor para agrupar esta información, con un índice de Silhouette de 0.72 donde se evidenció una clara separación de 10 o 15 años en los grupos, al comparar este rango de años de cuatro categorías con los cuatro grupos que agrupa k-means se encontró un 90% de exactitud, por lo que se concluyó con la predicción de que estos grupos pertenecen a temperaturas altas, muy altas, bajas y normales, según lo analizado en los mapas de calor, los gráficos de líneas e histograma de frecuencias. Se recomienda validar cada una de las técnicas para tener modelos efectivos que permitan tomar decisiones en el área de climatología.

Palabras clave: <ESTADÍSTICA>, <CLUSTERING>, <K-MEANS>, <MACHINE LEARNING>, <PYTHON (PROGRAMA)>, <PREDICCIÓN>, <MODELOS ATMOSFÉRICOS>.



15-12-2022

2397-DBRA-UPT-2022

ABSTRACT

The graduation work had the objective of predicting the behavior patterns of the temperature, where all the provinces of Ecuador were taken as the study population, there are few analysis dedicated to predicting behavior patterns, which is why the information obtained from the future atmospheric circulation models of the CHIRPS data bank was studied, this information had to be transformed into a two-dimensional matrix order to work with it, since it was in the form of a multidimensional matrix. When developing an exploratory analysis, an increasing trend was found that was not very pronounced yet clear, where it is evident that with the passage of time the temperature will have an increase, in addition its data come from a normal distribution, after applying several dimensionality reduction algorithms. As a result, it was found that the UMAP algorithm is better to find a two dimensional representation of this matrix with several columns in the present research, helping to preserve its multidimensional topological structure, likewise the Chebyshev distance was used to obtain better results, where the algorithm applied to make the k-means groupings was the best to group this information, with a Silhouette index of 0.72 where a clear separation of 10 or 15 years in the groups was evidenced, when comparing this range of years of four categories with the four sets that groups k-means, a 90% accuracy was found, so it was concluded with the prediction that these groups belong to high, very high, low and normal temperatures, as analyzed in the heat maps, line graphs and frequency histograms. It is recommended to validate each of the techniques to have effective models that allow decision making in the area of climatology.

Keywords: <STATISTICS>, <CLUSTERING>, <K-MEANS>, <MACHINE LEARNING>, <PYTHON (PROGRAM)>, <PREDICTION>, <ATMOSPHERIC MODELS>.



Edgar Mesias Jaramillo Moyano

0603497397

INTRODUCCIÓN

A lo largo de los últimos años el clima ha sufrido variaciones en la radiación solar recibida por el planeta tierra y estas han producido diferentes cambios en la temperatura. Mediante las arduas investigaciones de científicos expertos en la rama han podido corroborar que efectivamente existe incremento de la temperatura media global (Crowley & North, s.f., p. 996). El principal problema es la contaminación que en ciertas partes del mundo es más notable que en otras, así mismo, existen diversos factores que influyen en el cambio climático como son las emisiones de CO₂ emitidas por las grandes fábricas en la revolución industrial, siendo Ecuador un país pequeño que no emite mucho CO₂ a la atmosfera, sin embargo es el país que emite más emisiones que otros países Latino Americanos como Paraguay, Bolivia, El salvador, en consecuencia a medida que pasa el tiempo se observa que se va incrementando dichas emisiones según datos de (Crippa, et al., 2019, p. 97).

El Ecuador está constituido por 4 regiones: costa, sierra, oriente y la región insular, al ser una isla se considera el clima cálido al igual que la costa, la sierra posee un clima frío y el oriente al tener grandes selvas tiene un clima caliente y húmedo, además de esto existen los microclimas que según (Caballero, 2004, p. 1) .“En los niveles más bajos de la atmósfera, las diferentes características del diseño de espacios y el fenómeno de la isla de calor urbana ocasionan la individualización de áreas en función de pequeñas variaciones en las condiciones climáticas”.

Determinar el comportamiento de la temperatura temporal y espacial en muchos de los casos tiene un grado de complejidad. Es común el uso de datos teóricos que provienen de un modelo de pronóstico ya que representan aproximaciones que recopilan distintos fenómenos. Los modelos de circulación atmosférica nos dan pronósticos de lo que sucederá a futuro, el trabajo realizado intenta modelar patrones del clima que muy posiblemente sucederán debido a que estos modelos están calculados con miles de datos históricos lo que permite tener resultados con mayor precisión, además, que estos modelos consideran muchos factores físicos asegurando estimaciones muy precisas. Al contar con una serie temporal de los datos de los modelos de circulación atmosférica CMIP6, es preciso conocer los posibles resultados que se pueda obtener en esta investigación respecto a la afectación que producirá el cambio de temperatura, donde se podría observar los posibles escenarios, siendo muy útil para el sector agropecuario del país. La falta de información de temperatura en diversos sitios a largo plazo ha impedido obtener modelos con mayor precisión.

El Ecuador presenta las siguientes variaciones climáticas de modo que, en octubre la Cordillera Occidental representa la división climática, en el verano del hemisferio sur (diciembre), el sistema actual cambia tanto en la costa como en las tierras bajas amazónicas del Ecuador. En las tierras bajas orientales, los vientos desviados continúan extendiéndose hacia el hemisferio sur como vientos con presencia de lluvias (Bendix & Lauer, 1992, p. 133).

Los modelos de aprendizaje no supervisado nos permiten identificar grupos con características similares, ya sea con información bivariante e incluso multivariante, en donde desconocemos completamente lo que buscamos, simplemente este algoritmo nos ayuda a agrupar la información en donde cada grupo tengan alguna relación o sean semejantes para así poder categorizarlo de alguna forma mediante distancias. Existe también un problema a resolver con esta matriz de información que se obtiene, y es que los modelos de aprendizaje no supervisado no funcionan bien cuando trabajan con muchas dimensiones, por lo que es necesario reducir la dimensionalidad de la información para poder visualizar que estamos agrupando y no solo ver cuántos puntos pertenecen a cada grupo.

Debido a los cambios que se ha producido en la temperatura es importante investigar los diferentes patrones de comportamiento de la temperatura en el Ecuador.

Antecedentes

El comportamiento climático y su transición han provocado consecuencias relevantes en los seres humanos, afectaciones que contribuyen a pérdidas económicas en diferentes sectores. La elevación de la temperatura ha sido especialmente causada en las tres últimas décadas (1975-2005). Aproximadamente se han perdido ya más del 80% de los glaciares pirenaicos y para 2050 podrían desaparecer irreversiblemente (Nikolay, et al., 2010, p. 20). La variación de temperatura ha generado deshielo continental y el incremento de la temperatura del agua marina, provocando la subida del nivel del mar donde se prevé pérdidas especies. Desde el año 1998 se ha detectado tendencias crecientes, tanto en la temperatura media como en las temperaturas extremas en todo el Ecuador, se ha observado que la temperatura media presenta un cambio de 0.5°C a 1.6°C. Las temperaturas mínimas y máximas absolutas tienden a ser cada vez más altas, lo que evidencia los eventos extremos. La tendencia positiva en los valores de la temperatura mínima absoluta es más evidente en la región Interandina que en la región costanera, así mismo los incrementos más significativos tanto de temperatura mínima absoluta como de temperatura máxima absoluta, se producen en la estación de la región Sierra (Caceres, et al., 1998, p. 549).

En el análisis reciente de modelos climáticos para el país se muestra un incremento generalizado de la temperatura a nivel nacional; los incrementos tienen diferencias regionales, de acuerdo con el ensamble de los modelos climáticos globales seleccionados para el territorio nacional, utilizando el período de referencia 1981–2010, los resultados en los distintos escenarios muestran incrementos de la temperatura media anual (a nivel nacional) de 0.6 °C para el período 2011–2040, hasta 2.8 °C para fin de siglo, en el escenario optimista. La situación es análoga en cuanto a las temperaturas máximas, con aumentos de 0.8 a 3.5 °C, especialmente en la Amazonía y la Costa (Armenta, et al., 2016, p. 19)

Un reciente estudio realizado en el volcán Chimborazo determinó que en el período entre 1986 al 2013 la superficie del glaciar disminuyó en un 21% entre los principales impactos de los cambios del clima destacan aquellos relacionados a la disponibilidad de agua para las actividades productivas (agricultura y ganadería) y consumo de agua en las ciudades y comunidades de Ecuador (Armando, et al., 2012, p. 6). En el año 2016 en una investigación realizada mediante la efectuar de los modelos de CMIP 5 se evidencia que la temperatura máxima y mínima, al igual que la temperatura media, presentarían tendencias al aumento, con incrementos del orden de 0,8 a 3,5°C y de 0,6 a 2,8°C respectivamente. La temperatura máxima se vería reflejada en las regiones de la Amazonía y la Costa, en el caso de la temperatura mínima, la Amazonía sería la región con los aumentos más altos, del orden de 0,7-3,5°C a lo largo del siglo (Armenta, et al., 2016, p. 10).

PLANTEAMIENTO DEL PROBLEMA

Enunciado del problema

Es factible deducir que el sistema climático de Ecuador es vulnerable ante diversas dinámicas que se presentan actualmente, o a futuro, entre las que se cita a las potenciales alteraciones de las corrientes marinas, y los cambios en los patrones de humedad provenientes de la Amazonía, todo lo cual podría incrementar la vulnerabilidad de las poblaciones (Buytaert & De Bievre, 2012, p. 10).

La degradación del medioambiente se relaciona con escasez de agua, contaminación de los océanos, minería, deforestación, sequías prolongadas y glaciares derretidos que infieren en la desaparición de especies, a esto se suma además la deforestación, el calentamiento global y el cambio climático.

Entre los principales impactos de los cambios del clima es la reducción de los glaciares que preocupa a la población ya que es útil e indispensable para el consumo humano, agricultura y ganadería, en el sector agrícola el cambio climático provoca la propagación de hongos en los sembríos, la desaparición de especies ha acontecido en los últimos años aproximadamente 667 especies (angiospermas, helechos, anfibios, reptiles y aves), se encuentran en peligro de extinción debido a la modificación de su hábitat todo esto generado por el cambio drástico del tiempo (Armando, et al., 2012, p. 6)

No se han encontrado estudios similares en el Ecuador, por lo tanto, esta investigación aporta a la predicción de eventos extremos (lluvias y sequías) que se podría mitigar los efectos que alteran los procesos de producción en la agricultura, propagación de enfermedades, fauna y flora.

Formulación (Incógnita)

¿Cuáles son los patrones de comportamiento de la temperatura de los años 2020-2050?

¿Cuál es la mejor técnica de Machine Learning que ha producido mejores resultados para determinar patrones de comportamiento?

¿Qué tan preciso es la técnica Machine Learning para alcanzar el objetivo de la investigación?

JUSTIFICACIÓN

La presente investigación es de gran relevancia por diversas razones, como el hecho de que el clima es muy variante a lo largo del tiempo a ello supone retos para la humanidad ya que un cambio abrupto eleva el nivel del mar y da lugar a inundaciones, y esto genera un riesgo para especies en peligro de extinción.

Otra de las razones es que se necesita un modelo que nos ayude a determinar el periodo más frío y seco con el fin de aportar con información a los agricultores y puedan tomar decisiones oportunas para evitar grandiosas pérdidas económicas. En el Ecuador no existen diversas investigaciones enfocadas a la aplicación de técnicas Machine Learning que permitan agrupar datos por características similares en relación con variables meteorológicas, lo que viene a ser algo nuevo e intrigante saber el resultado del comportamiento de la temperatura futura para así averiguar los climas más característicos que tiene el Ecuador.

La falta de estudios relacionados a los patrones de comportamiento motiva a tener como prioridad investigar la conducta a posteriori del clima en el Ecuador donde contribuyan a la creación de nuevos modelos de predicción en el que la información será de gran utilidad para la toma de decisiones y erradicar los problemas que se generen por el incremento y disminución acelerada de la temperatura.

OBJETIVOS

Objetivo general

Predecir patrones de comportamiento de la temperatura mediante herramientas estadísticas usadas en los modelos de circulación atmosférica de 2020 a 2050.

Objetivos específicos

- Explorar los datos obtenidos de los modelos atmosféricos y estaciones meteorológicas del INHAMI.
- Predecir patrones de comportamiento de temperatura mediante técnicas Machine Learning.
- Validar el modelo mediante comparación de datos observados y datos predichos por el modelo.

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL

1.1. Bases conceptuales

1.1.1. Estadística

La Estadística es una ciencia que tiene como finalidad obtener, recopilar, procesar, resumir y presentar datos respecto a un estudio de interés, transformándolos en estadísticas y en base a ello realizar interpretaciones estadísticas para obtener conclusiones y tomar decisiones, brindando seguridad de idoneidad en los procedimientos (Rustom, 2012, p. 9).

1.1.1.1. Estadística descriptiva

Se encarga de la recopilación, clasificación, presentación y descripción de los datos mediante tablas o gráficas, y en base a ello el cálculo de parámetros básicos (García, 2011, p. 11).

1.1.2. Medidas tendencia central

Permiten condensar, en un único valor, los resultados obtenidos para la totalidad y en relación con cada una de las variables consideradas. Lo que pretende en general es observar el comportamiento mayoritario de la muestra o población (García, 2005, p. 135).

1.1.2.1. Media aritmética

Media o promedio se lo define como al conjunto de (N,n) mediciones y su cálculo esta atribuido a la suma de sus valores dividido entre (N,n) donde : La media poblacional es representada por (μ) y la media muestral con \bar{x} , las fórmulas vienen expresadas de la siguiente manera:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

1.1.2.2. Mediana

Es el valor de la variable que ocupa la posición central o las dos posiciones centrales de los datos ordenados, es decir son valores que aparta a los datos ordenados en dos grupos con el mismo

número de observaciones, el primero con valores mayores o iguales a la mediana y el segundo con valores menores o iguales a la mediana (Rustom, 2012, p. 17).

$$\tilde{x} = L_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i}$$

1.1.3. Desviación estándar

Mediante la desviación estándar se puede tener una perspectiva de referencia con el fin de estimar la variación que ocurre en un proceso, viene definido como la raíz cuadrada de la varianza, cuya expresión es (Rustom, 2012, p. 19).

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$

1.1.4. Varianza

Su valor pertenece a la desviación cuadrática media de los datos de la media aritmética. En otras palabras, su valor da una idea de cuánto varía el dato con respecto a la media aritmética, la variancia para un conjunto de datos viene dada en unidades cuadráticas (Matos & Fortunato, 2020, p. 128).

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}$$

1.1.5. Gráficas Estadísticas

Las gráficas estadísticas ayudan al lector a comprender lo que sucede con la información recopilada, la principal ventaja es la observación que existe entre la relación de los datos.

1.1.5.1. Histogramas

El histograma se elabora colocando la variable de interés en el eje horizontal y la frecuencia % o la frecuencia relativa en el eje vertical. Las frecuencias de cada clase se indica dibujando un

rectángulo cuya base está definida por los límites de clase sobre el eje horizontal y cuya altura es la frecuencia (Ray, 2008, p. 36).

1.1.5.2. Diagrama de puntos

Es una forma de resumir datos cuantitativos, donde las observaciones individuales son representados por un punto sobre una recta numérica, cuando se posee varios datos, cada uno de los puntos puede representar un número fijo de individuos, una de las ventajas es que se puede verificar si existen observaciones inusuales (Galindo, 2011, p. 10).

1.1.6. Aprendizaje no supervisado

En el ANS, no existe un supervisor y se identifica además por tener datos de entrada, el propósito es hallar las regularidades en la entrada. Hay una estructura en el espacio de entrada tal que ciertos patrones ocurren con más frecuencia que otros, y se requiere ver qué sucede generalmente y qué no. En estadística, esto se llama estimación de densidad. Un método para la estimación de la densidad es el agrupamiento, donde el objetivo es encontrar agrupaciones de entrada (Alpaydm, 2014, p. 11).

1.1.7. Reducción de dimensionalidad

La reducción de dimensionalidad es utilizada con mayor frecuencia en el procesamiento de particularidades con la finalidad de seleccionar datos de un espacio dimensional alto y mapearlo en un subespacio dimensional más pequeño, pero manteniendo la mayor parte de la información importante (Mirjalili & Raschka, 2017, p. 10).

1.1.7.1. Componentes principales

Es una técnica que permite obtener nuevos atributos como combinación lineal de factores independientes respecto a la limitación o reducción del número de atributos iniciales (García & Molina, 2018, p. 195).

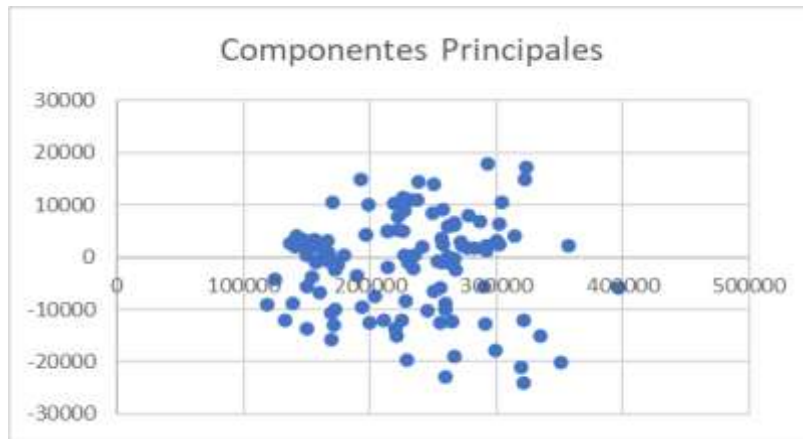


Gráfico 1-1. Componentes Principales

Realizado por: Toapanta, Erika, 2022.

1.1.7.2. Incrustación de vecinos estocásticos distribuidos en T

Es un algoritmo que permite reducir la dimensionalidad atribuidos en la similitud de puntos en base a características semejantes, además se encarga de salvaguardar la información donde muestra los puntos vecinos entre sí (Müller & Guido, 2016, p. 164) .

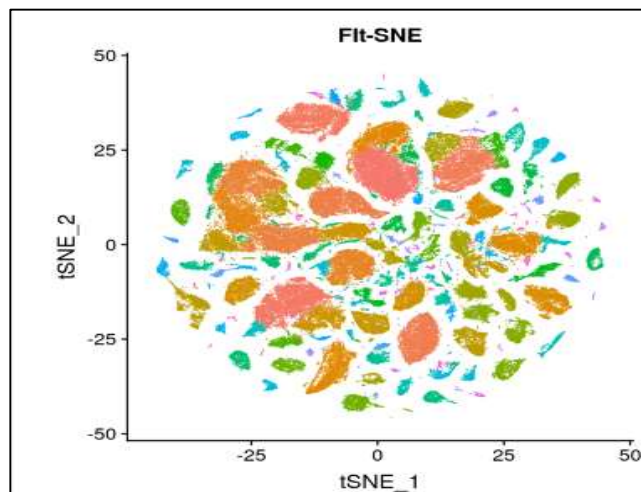


Gráfico 2-1. Algoritmo TSNE

Fuente: ICHIPRO, 2022.

1.1.7.3. Aproximación y proyección uniforme del colector

UMAP es otro de los algoritmos que permite realizar reducción de dimensionalidad no lineal, se encarga de crear representaciones topológicas de los datos en relación de la dimensión original donde tiene como objetivo aproximar los datos y en dimensión baja se inicia con valores totalmente aleatorizados.

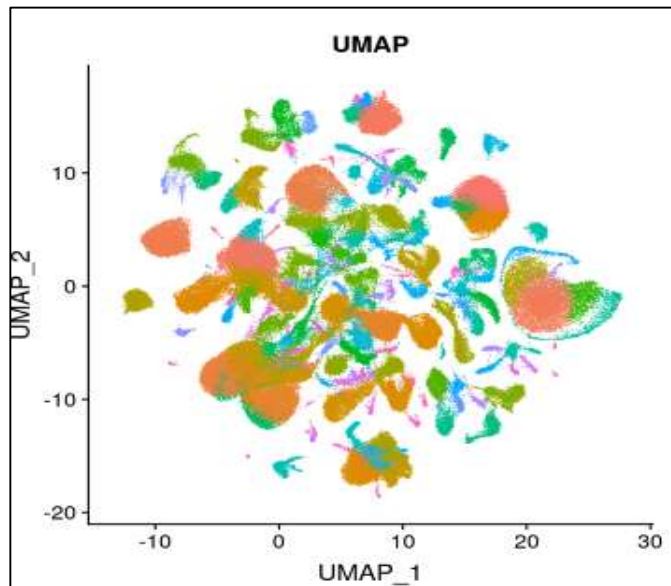


Gráfico 3-1. Algoritmo UMAP

Fuente: ICHI.PRO, 2022.

1.1.8. Agrupamiento o clustering

Es una técnica exploratoria de análisis de datos que nos permite organizar numerosa información en subgrupos denominado clúster, sin antes haber tenido algún conocimiento previo de los miembros del grupo. Cada uno de los clústeres que surgen durante el análisis define un grupo de objetos que comparten similitud en un cierto grado, pero difieren de los objetos de otros clústeres, razón por la cual el agrupamiento también es denominado a veces clasificación sin supervisión. El agrupamiento es una excelente técnica para estructurar información y derivar relaciones significativas de los datos (Mirjalili & Raschka, 2017, p. 7).

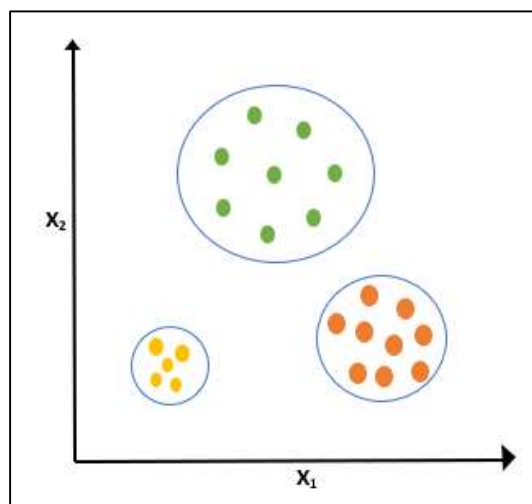


Gráfico 4-1. Agrupamiento de datos en 3 grupos diferentes.

Realizado por: Toapanta, Erika, 2021.

1.1.9. K-means

Es una técnica de agrupamiento particional simple basada en prototipos que intenta encontrar los agrupamientos K especificados por el usuario. Estos grupos están representados por sus centroides que pueden ser la media aritmética o la moda de los objetos de datos en un cierto grupo según se especifique sus atributos de manera numérica o simbólica, así mismo en base a diferentes funciones de distancias pueden conducir a diferentes tipos de K-medias (Wu, 2012, p. 71).

1.1.10. Distancias

Diversos criterios geométricos y en la noción de distancia entre poblaciones e individuos están basados en algunos métodos de AM.

1.1.10.1. Distancia Euclídea

Supone implícitamente que las variables están correlacionadas y no es invariante para cambios de escala (Cuadras, 2018, p. 20).

$$d_E(i, j) = \sqrt{\sum_{h=1}^p (x_{ih} - x_{jh})^2}$$

1.1.10.2. Distancia Chebyshev

Es una distancia definida en un espacio vectorial en donde la distancia entre dos vectores es el mayor de sus diferencias a lo largo de cualquier dimensión de coordenadas (Rodríguez, 2015, p. 9)

$$d_{chev}(i, j) = \max_k |x_{ik} - x_{jk}|$$

1.1.11. Clustering Jerárquico

Los métodos de agrupamiento jerárquicos construyen estructuras en forma de árbol, o llamadas dendrogramas, así mismo los elementos se agrupan o separan para minimizar una medida de distancia o maximizar una medida de similaridad. En este sentido, podemos clasificar los métodos jerárquicos en:

- Aglomerativos. Se parte de tantos clústeres como elementos hay y se van agrupando en subconjuntos cada vez mayores hasta llegar a un único clúster.
- Disociativos. Se parte de un único clúster que contiene todos los elementos, que se va dividiendo en subconjuntos cada vez más pequeños (Aguilera, et al., 2019, p. 12).

1.1.12. Índice Silhouette

El cálculo de los índices es necesario cuando las proximidades están en una escala de razón por ejemplo en las distancias euclidianas así mismo cuando necesitamos conocer agrupaciones separadas y compactas, esta herramienta estadística permite validar un resultado de aplicación de clúster o conglomerado. Para la elaboración del índice es necesario presentar la partición obtenida esta extraída de cualquier técnica de agrupamiento y la colección de todas las proximidades entre objetos (Rousseeuw, 1987, p. 55).

1.1.13. Matriz de confusión

La matriz de confusión permite evaluar el rendimiento de modelos, es una de las herramientas más empleadas en Machine learning, “confusión matriz” es una tabla con filas y columnas que contabilizan las predicciones en comparación con los valores reales. Mediante la matriz empleada se obtiene una mejor comprensión en el comportamiento del modelo si lo hace bien o mal, así mismo es muy útil para mostrar de forma explícita cuando una clase es confundida con otra (Torres, 2020, p. 50).

		PREDICCIÓN	
		POSITIVOS	NEGATIVOS
Observacione	POSITIVOS	Verdaderos Positivos VP	Falsos Negativos FN
	NEGATIVOS	Falsos Positivos FP	Verdaderos negativos VN

Figura 1-1. Matriz de confusión para una clasificación binaria

Fuente: Torres Jordi, 2005.

1.1.14. Almacenamiento de datos NetCDF

La estructura de datos en los datos netCDF viene dado por cada conjunto que contiene dimensiones, variables y atributos, cada uno de los cuales son identificado por una jerarquía de nombres únicos. el conjunto de datos y los objetos variables pueden contener atributos que describen el contenido, las unidades, la historia u otros metadatos del objeto (Hoyer & Hamman, 2017, p. 2).

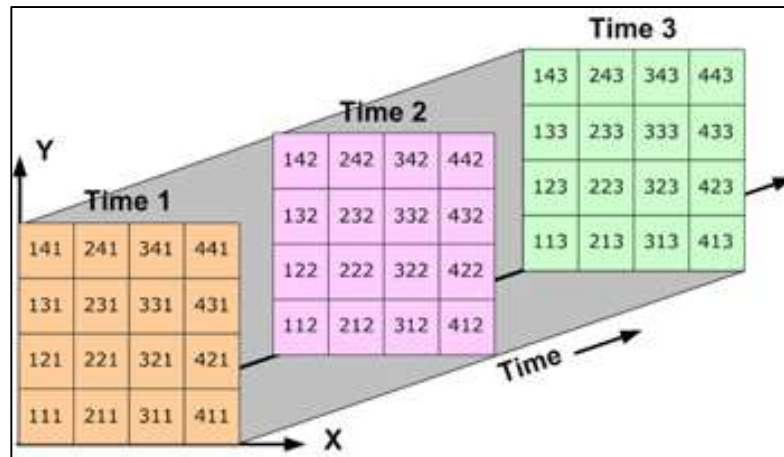


Figura 2-1. Datos de un área que varían con el tiempo

Fuente: ARGIS, 2019

1.1.15. Dataset

Xarray.Dataset es el equivalente multidimensional de xarray de un DataFrame. El contenedor es muy parecido a un dict de matrices etiquetadas u denominados objetos DataArray donde sus dimensiones vienen alineadas. Está diseñado como una representación en memoria del modelo de datos del formato de archivo netCDF. Cabe mencionar que la interfaz es similar a un dictado del propio conjunto de datos, que se puede emplear para ingresar a cualquier variable en un conjunto de datos. En la gráfica propuesta se puede llamar a la temperatura y la precipitación "variables de datos" y + las demás matrices "variables de coordenadas" ya que etiquetan los puntos a lo largo de las dimensiones (Hoyer & Hamman, 2017, p. 3)

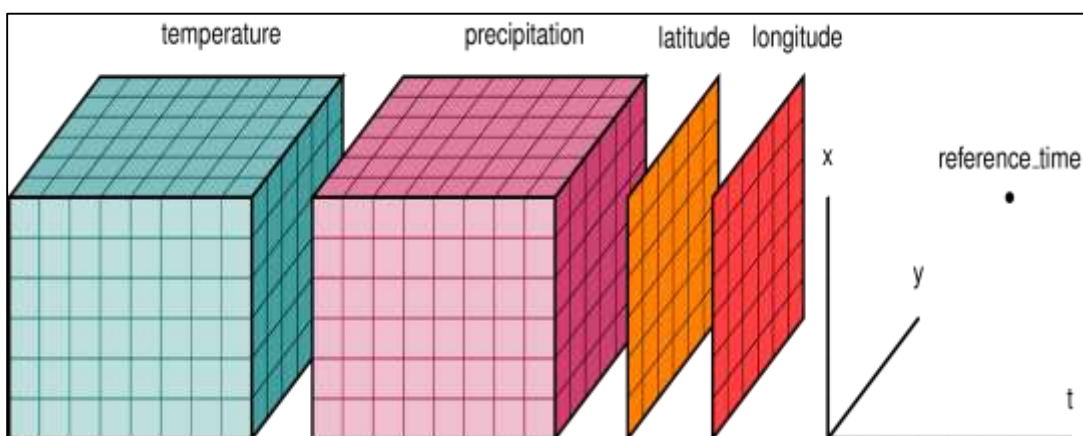


Figura 3-1. Estructura de un Dataset del paquete xarray

Fuente: Hoyer, S; Hamman, J, 2021

1.1.16. Anaconda

Es una herramienta fundamental en la ciencia de datos cabe destacar que es de código abierto, sus paquetes son varios aproximadamente constan de 100 y más basados en Python, Scala y R, además posee un administrador de paquetes denominado conda donde el usuario alcanza a trabajar con cientos de paquetes donde la diversidad de idiomas es variada de manera que es una plataforma completa donde se efectúa el preprocesamiento de datos, validación, agrupamiento y clasificación de modelos (Yan & Yan, 2018, p. 1).

1.1.17. Jupyter

Es una interfaz de código abierto su entorno computacional es interactivo donde su principal objetivo es realizar análisis de ciencia de datos establecidos por Python, el entorno interactivo es ideal para el desarrollo y nos permite compartir de manera sencilla el cuaderno donde se trabaja (Sarkar, et al., 2017, p. 72).

1.2. Bases teóricas

1.2.1. Cambio climático

Los cambios climáticos son variaciones a largo plazo en los valores medios de cierta variable climática, donde los valores medios vienen correspondidos a las fluctuaciones estacionales y anuales, donde su influencia puede ser dada por el cambio en el valor de su media, basados en la escala de tiempo humana, los cambios en promedio son tan lentos que son prácticamente paulatinos donde su detección solo se lo puede hacer mediante un registro minucioso de manera que no siempre se mantendrá o será repetitivo por las diferentes causas naturales que estas provocan (Casas & Martha, 1999, p. 118).

1.2.2. Meteorología

Se encarga del estudio del estado del tiempo, los fenómenos, su medio atmosférico y las leyes que lo rigen su campo proviene de la física atmosférica. El estado del tiempo meteorológico atribuye a los conjuntos de particularidades que muestra la atmosfera en un lugar o momento específico como la temperatura, humedad, viento, precipitación, presión, nubosidad; estas variables, sus interacciones y sus manifestaciones tienen lugar en la tropósfera, bajo la estratosfera, finalmente la meteorología se encarga de evaluar el estado del tiempo en un momento actual, en base a ello intenta predecirlo en un futuro cercano (Lozano, 2018, p. 23).

1.2.3. Componentes del sistema climático

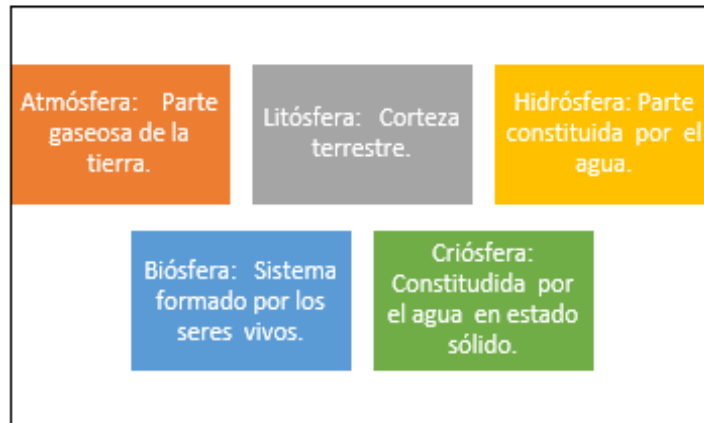


Figura 4-1. Componentes del Sistema climático

Fuente: Casas Tomas, 1997

1.2.4. Topografía y geología

El Ecuador continental está situado al Noroeste de América del Sur, entre los 01° 28' de Latitud Norte y 05° 01 de Latitud Sur y desde los 75°11 en la planicie Amazónica hasta los 81° 01 de longitud Oeste. Limita al norte con Colombia, al sur y al este con Perú y al oeste con el océano Pacífico. El territorio del Ecuador continental está dividido en tres regiones naturales claramente diferenciadas entre sí, ya sea por su topografía, clima, vegetación y población (Farfán, 2018, p. 18).

1.2.4.1. Costa

Es una zona de llanuras fértiles, colinas, cuencas sedimentarias y elevaciones bajas, compuesta por 7 provincias, que tienen un clima tropical o ecuatorial (Farfán, 2018, p. 15).

1.2.4.2. Sierra

La Región Sierra se encuentra dominada por la cordillera de los Andes, cruza por el Ecuador de norte a sur formando dos cadenas montañosas, la occidental y la oriental. La diversidad en el clima lo caracteriza, debido a la presencia de la cordillera de los Andes y a los vientos que soplan por los valles y llanuras, es una de las regiones con mayor número de provincias en su totalidad 10 (Farfán, 2018, p. 16).

1.2.4.3. Amazonía

En la Amazonía se encuentran seis provincias del Ecuador, el relieve de la Amazonía o región oriental está conformado por colinas que se originan en la parte oriental de los Andes y descienden hasta las llanuras del Amazonas, a cuya cuenca pertenecen importantes ríos (Farfán, 2018, p. 12) .

1.2.4.4. Región Insular

La región insular o islas Galápagos está constituido por 13 islas mayores, 6 menores y 42 islotes de origen volcánico que suman una extensión superficial de 8 010 kilómetros cuadrados el océano Pacífico a 1 120 kilómetros de la costa nacional Temperaturas medias de las cuatro regiones del Ecuador.

1.2.5. Clima

Se caracteriza por ser un conjunto de valores promedio de las condiciones atmosféricas como: precipitación, temperatura, presión, humedad que caracterizan a una zona cabe mencionar que su variación se da por las diferentes regiones y a través del tiempo (Federal, 2009, p. 13).

1.2.6. Microclima

Es una condición climática que se destaca en sus características del clima de su entorno, por diferentes elementos climáticos y geográficos específicos un lugar. Los acontecimientos se dan por diversas razones naturales, como por ejemplo la orientación de una montaña, que genera microclimas más húmedos o cálidos, bajo un bosque primario, o también en las grandes ciudades, donde se concentra el calor y se reduce el viento por efecto de la emisión de gases contaminantes (Viguera & Ruth, 2017, p. 8).

1.2.7. Eventos extremos

Se produce un cambio climático debido al incremento en la frecuencia o intensidad de eventos climáticos extremos que vienen acompañadas por sequías fuertes que afecta a la agricultura del Ecuador y precipitaciones rápidas (Riebeck, 2005, p. 1).

Se dice que ocurre un evento extremo cuando está presente en un área y puede ser bastante normal en otra, y parece claro que la naturaleza y la sociedad están más adaptadas a sus promedios de lo que están, en base a ello se puede tener en cuenta las frecuencias que son ocurridas en un determinado lugar o tiempo (Serrano.Sheila, 2012, p. 26).

1.2.8. Circulación general de la atmosfera

La circulación atmosférica está compuesta por un sistema de vientos y presiones a escala planetaria como consecuencia de la diferente radiación solar recibida, al efecto de la rotación terrestre y al efecto de la continentalidad, los movimientos de sistemas meteorológicos de la escala más diminuta son atribuida por la circulación planetaria (Nadal & Muñuzuri, 2006, p. 106).

1.2.8.1. Humedad

La humedad atmosférica está constituida por gotas de agua, vapor y las nubes formadas por cristales de hielo, el contenido de humedad se encuentra determinado por la temperatura del aire, evaporación local y el transporte atmosférico horizontal de humedad. En promedio el agua de las nubes, escala a sólo el 4% de la humedad atmosférica (Roger & Chorley, 2003, p. 83).

1.2.8.2. Densidad del aire

La densidad del aire es la absorción limitada que produce un gran aumento de temperatura. Las temperaturas aumentan bastante generalmente con altura en verano, con los más fríos aire en la tropopausa ecuatorial. En invierno, la estructura es más complejo con temperaturas muy bajas (Roger & Chorley, 2003, p. 27).

1.2.9. Temperatura

La temperatura del aire normalmente utilizada en climatología es la registrada en la superficie. humedad o la falta de humedad modifica la temperatura. Cuanta más humedad haya en una región, menor será el rango de temperatura, y cuanto más seca la región, mayor el rango de temperatura. La humedad es también influenciada por la temperatura. El aire más cálido puede contener más humedad que un aire más frío, lo que resulta en una mayor evaporación y probabilidad de nubes y precipitaciones. (Gaurishankar, 2012, p. 273).

1.2.10. Variabilidad natural

Depende de las variaciones en la energía solar externa forzando, así mismo de dos características del propio sistema: retroalimentación y comportamiento no lineal. El sistema climático se deriva de los cambios en la atmósfera, composición resultante de la acción humana induciendo al

aumento de los gases de efecto invernadero, que en temperaturas globales producen un cambio e incremento en material como: carbono y polvo mineral, aerosoles. Además, las partículas, aerosoles volcánicos, que ingresan a la estratosfera, influyen en el clima global (Roger & Chorley, 2003, p. 7).

1.2.11. Radiación terrestre

La temperatura media de la Tierra se encuentra aproximadamente en 288 K, además emite radiación como cualquier otro cuerpo. El resultado de este ajuste es que la Tierra radia al exterior como un cuerpo negro a la temperatura de 255 K. Temperatura claramente inferior a la de la superficie terrestre como consecuencia del llamado efecto invernadero (Zuñiga & Crespo, 2015, p. 35).

1.2.12. Efecto invernadero

Se le denomina al incremento de la temperatura de la superficie terrestre, atribuido al calentamiento de la atmósfera y esto produce radiación, así mismo el efecto invernadero ha generado problemas significativos en la actualidad en los últimos 50 años producido por el ser humano generando un incremento acelerado en la temperatura media (Zuñiga & Crespo, 2015, p. 42).

1.2.13. Variaciones de temperatura

- La variación de la temperatura puede ser generada por la cantidad de energía solar recibida, en las diferentes regiones del planeta conforme a su variación por la hora del día, con la estación del año y latitud, las diferencias de radiación solar provocan una variación en la temperatura, así mismo se puede dar debido a la distribución de distintos tipos de superficies y en función de la altura.
- Afectan la temperatura los siguientes factores como el: cambio de día, distribución de latitud, cambio estacional, tipos de superficie terrestre, cambio de altitud, cabe destacar que la variación diurna se da por el cambio de temperatura entre el día y la noche debido a la rotación de la Tierra. (Sarochar, 2019, p. 5).

1.2.14. Anomalía climática

Se construyen índices mensuales y de periodo extendido que detectan anomalías o alteraciones del comportamiento de la temperatura del aire, con respecto a la media multianual (Montealegre Bocanegra José, 2007, p. 30).

1.2.15. Temperatura media global

La temperatura media global absoluta, ha sido estimada desde los años 1951-1980 con una temperatura media global del aire en la superficie como 14°C con una incertidumbre de varias décimas de grado Celsius, el valor se obtuvo mediante el uso de un modelo climático global (Hansen & Ruedy, 2010, p. 3).

CAPÍTULO II

2. MARCO METODOLÓGICO

2.1. Tipo de investigación

Respecto con el método de investigación es de tipo cuantitativo que se caracteriza por presentar una secuencia de procesos, que permiten llegar al objetivo de la investigación.

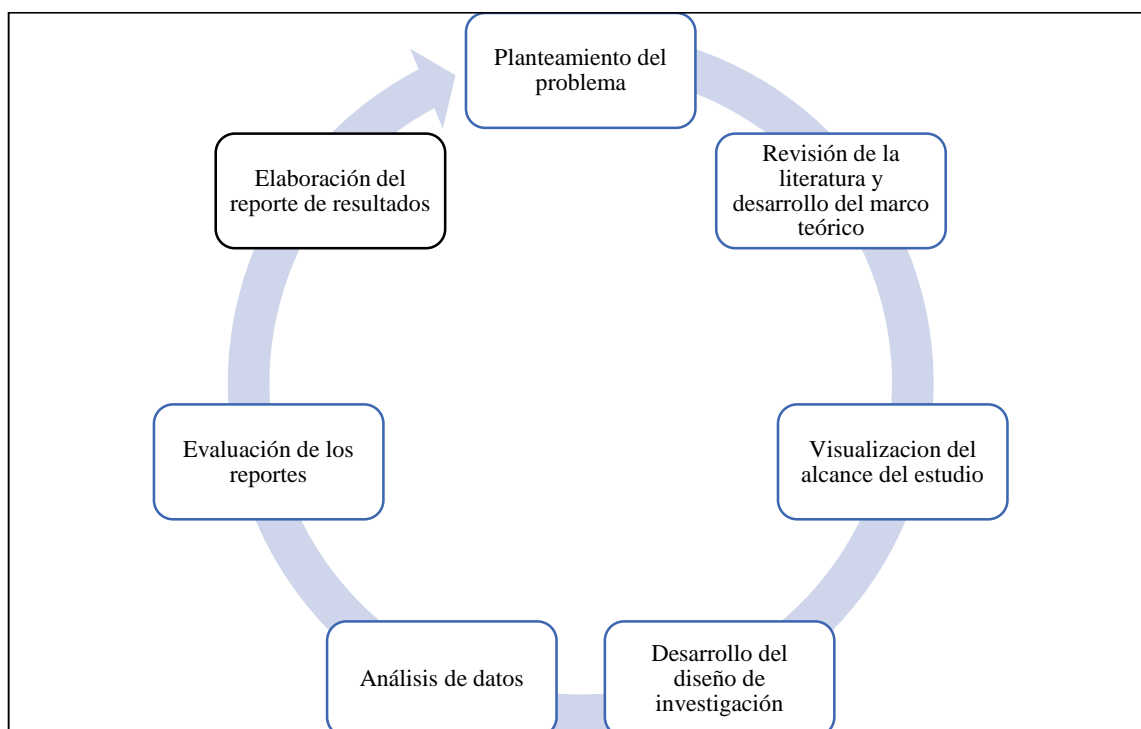


Figura 1-2. Proceso cuantitativo

Fuente: (Hernández & Fernandez, 2016, p. 5)

Realizado por: Toapanta, Erika, 2022.

2.1.1. Planteamiento del problema

La variabilidad del cambio climático ha desatado grandes consecuencias en la agricultura provocando sequías, fuertes lluvias y aún más preocupante el deshielo en los glaciares del Ecuador, por ende, es de vital importancia estudiar, descubrir y analizar los patrones meteorológicos del país.

2.1.2. Revisión de la literatura y desarrollo del marco teórico

En el Ecuador no existe estudios que investiguen el comportamiento climático, se ha encontrado 2 artículos científicos respecto a las técnicas machine learning donde se describe las diferentes técnicas utilizadas para determinar pronósticos. En el área de climatología las técnicas de machine learning son muy utilizadas, en la literatura científica se ha encontrado aproximadamente mil artículos relacionados, de los cuales se ha considerado las 40 investigaciones más relevantes que ayudarán a contrastar con los resultados obtenidos. Se ha considerado fuentes primarias como lo es 20 artículos científicos relacionados a la climatología en el mundo y el Ecuador; fuentes secundarias: 2 libros metodología investigación, 9 Python , 4 análisis multivariante y fuentes terciarias: páginas web para encontrar códigos en el lenguaje de programación como Atmospheric Data Community Toolkit.

2.1.3. Visualización del alcance del estudio

El alcance de esta investigación es exploratorio, debido que propone descubrir los patrones de comportamiento climático futuro en el Ecuador, así mismo es descriptiva, de manera que se espera representar como se manifiestan los diferentes patrones de comportamiento encontrados según el tiempo.

2.1.4. Desarrollo del diseño de investigación

El diseño de investigación se la define como no experimental ya que se busca conocer los patrones de comportamiento de la temperatura que se presentará en los transcurso de los años 2020-2050, según al periodo es longitudinal o evolutiva ya que se estudian datos de la temperatura ambiental de un determinado periodo donde se va a examinar el cambio a través del tiempo.

2.1.5. Análisis de datos

Se ha desarrollado un análisis crítico de los resultados contrastados con la bibliografía mediante gráficos estadísticos como: histogramas de frecuencias, gráficos de líneas, de igual manera se aplicó inferencia utilizando técnicas multivariantes. Mediante el lenguaje de programación estadístico Python se efectuó modelos de reducción de dimensionalidad, PCA, TSNE y UMAP ,posteriormente se efectuó un gráfico de codo para conocer el número de clústeres o grupos con los que se va a trabajar , seguidamente se comprobó los diferentes algoritmos para el agrupamiento de datos como K-means, DBSCAN, Agglomerative Clustering con la finalidad de conocer el comportamiento de los patrones de temperatura mediante la agrupación de

características similares, finalmente se validó los modelos mediante la aplicación de diferentes métricas y gráficas .

2.1.6. Evaluación de los reportes

Se ha calculado el número de clústeres mediante la gráfica de codo, así mismo la confiabilidad de los algoritmos ha sido evaluada mediante métricas de precisión como: Siluete colectivamente con sus diferentes métricas.

2.1.7. Elaboración del reporte de resultados

Se ha realizado un análisis de los resultados contrastados con la bibliográfica empleada mediante las interpretaciones de las gráficas y métricas estadísticas.

2.2. Diseño de la Investigación

Se utilizó el método de investigación cuantitativa y según la manipulación de variables es un diseño no experimental ya que la información procede de fuentes secundarias del grupo de investigación GEA.

2.2.1. Localización del estudio

El estudio está focalizado en todo el país ecuatoriano, los datos obtenidos son de cada sector del Ecuador cubriendo todo el país.



Figura 2-2. Mapa del Ecuador

Fuente: Google maps, 2021

2.2.2. Población de estudio

Anomalía de la temperatura en todo el país ecuatoriano con una frecuencia diaria.

2.2.3. Tamaño de la muestra

No se toma una muestra, ya que se trabaja con el total de observaciones pronosticadas desde el año 2020 hasta el año 2050 correspondiente a CMIP6

2.2.4. Método de muestreo

No se tiene ningún método de muestreo, ya que no se tiene ninguna muestra.

2.2.5. Técnica de recolección de los datos

Los datos CMIP6 fueron obtenidos de la Biblioteca de Datos UCSB CHIRTS v2p0 daily global 0p05 prcp disponible en: <http://iridl.ldeo.columbia.edu/SOURCES/.UCSB/.CHIRPS/.v2p0/.daily/.global/.0p05/T/2457356.0/VALUE/.prcp/>

2.2.6. Modelo estadístico

Los modelos que se probarán para estos datos tienen que ver con modelos de aprendizaje no supervisado que son conocidos como modelos de agrupación o de conglomerados ya que nos permite aglomerar los datos con características similares, así mismo la validación de los resultados mediante métricas estadísticas.

2.2.7. Operacionalización de las variables

Tabla 1-2: Operacionalización de variables

Nombre de la variable	Descripción	Tipo de variable	Escala de medición	Categoría o intervalo
TEMPERATURA	Temperatura del clima en el Ecuador	Cuantitativa Continua	Intervalo	$(-\infty, +\infty)$

Realizado por: Toapanta, Erika, 2022.

CAPÍTULO III

3. MARCO DE RESULTADOS Y DISCUSIÓN DE LOS RESULTADOS

A partir de la información de la variable temperatura se procede a realizar la exploración de los datos, posteriormente son empleadas las técnicas de reducción de dimensión por consiguiente se aplica las técnicas machine learning para identificar patrones de comportamiento, finalmente se efectúa la validación del modelo con métricas estadísticas y gráficas.

2.1. Descripción de la matriz de información

Se observa que la variable temperatura posee tres dimensiones distintas con sus respectivas coordenadas definidas por: Y latitud, X longitud y T es tiempo asimismo los datos que han sido recolectados a escala diaria.

Tabla 1-3: Tabla de la variable Temperatura bajo sus tres dimensiones

Dimensiones (T: 10954, X: 34, Y: 22)					
Coordenadas:					
*Y	(Y) float	-16.11	-11.91	9.105	13.31
*X	(X) float	-10.5	-10.2	-61.88	-59.06
*T	(T) datetime64[ns]	2020-01- 03T12:00:00		2049-12- 29T12:00:00	
Variable de datos					
adif (T, Y, X) float32 ...>					
Realizado por: Toapanta, Erika, 2022.					

La base de datos de la temperatura procede de los modelos atmosféricos del Centro Nacional de Investigaciones Meteorológicas (CNRM) que se puede encontrar en el proyecto CMIP6, son datos correspondientes al periodo 2020-2050, cuenta con un formato NetCDF el cual almacena datos multidimensionales como las coordenadas: latitud, longitud y tiempo. Los datos fueron recolectados a escala diaria de 137 estaciones meteorológicas del INAMHI.

3.1. Estadísticas descriptivas

Para poder trabajar con una matriz bidimensional extensa se debe obtener un promedio ya sea por fila o columna para pretender describir lo que está sucediendo, a continuación, se realizará una reducción de la dimensionalidad para conseguir otro enfoque con todas las variables.

Tabla 2-3: Estadísticos calculados por promedio de filas.

Estadístico	Resultado
Total, de observaciones	10954
Media	0.000009
Desviación estándar	0.700250
Mínimo	-1.534581
Cuartil 25%	-0.342224
Mediana	-0.014340
Cuartil 75%	0.322334
Máximo	1.761686

Realizado por: Toapanta, Erika, 2022.

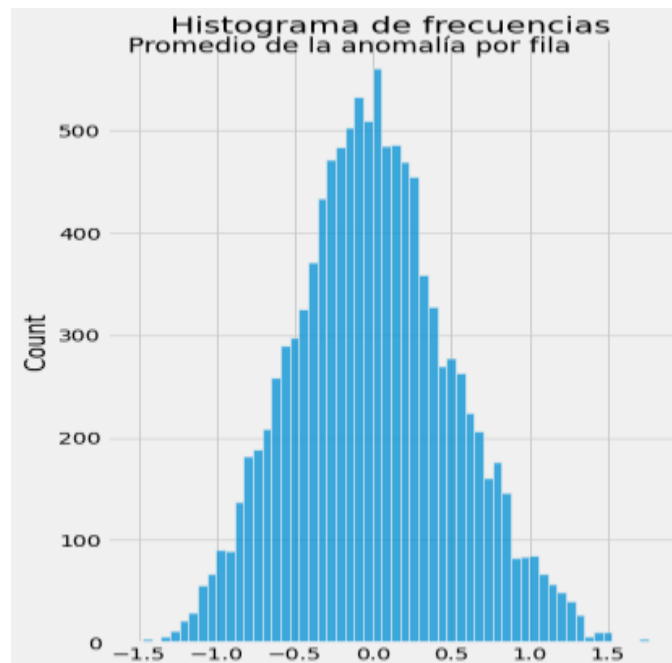


Gráfico 1-3. Histograma de frecuencias de las anomalías.

Realizado por: Toapanta, Erika, 2022.

En el histograma de frecuencias se observa que la variable temperatura siguen una Distribución de Probabilidad Normal con media cero y varianza 1.

3.1.1. Exploración de la variable temperatura

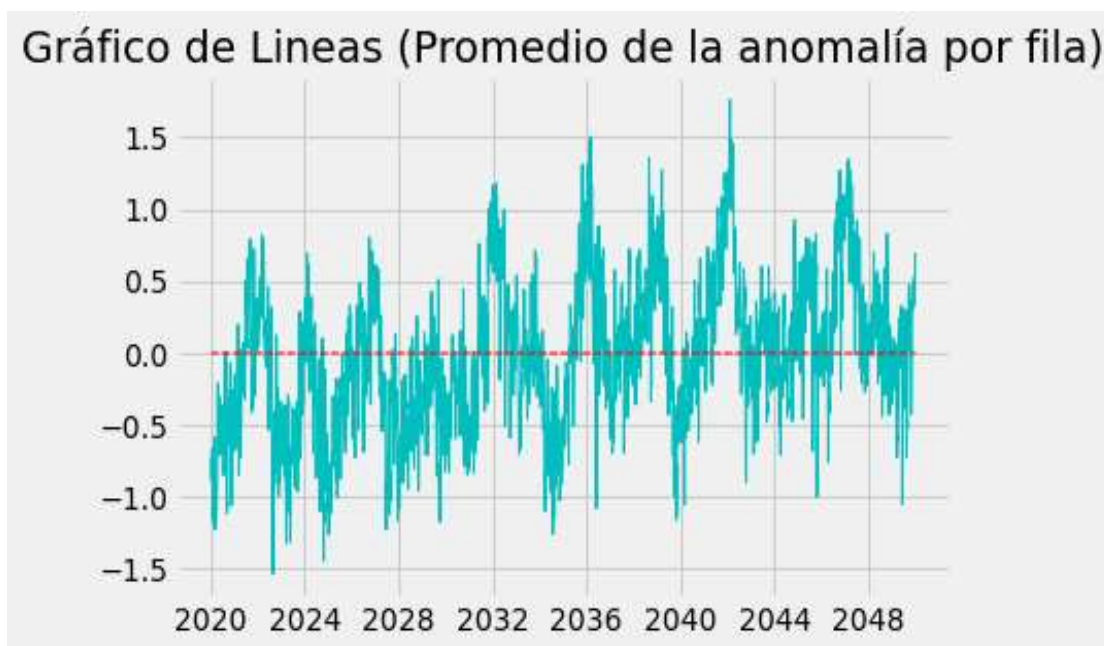


Gráfico 2-3. Gráfico de líneas del promedio de filas por año.

Realizado por: Toapanta, Erika, 2022.

El **Gráfico 2-3**, se observa una tendencia creciente de la variable temperatura a medida que pasa el tiempo, esto implica una consecuencia negativa a futuro que puede afectar a la biodiversidad de nuestro país.

3.1.2. Matriz bidimensional

La presente matriz posee los datos apilados para poder efectuar clustering y tiene las siguientes dimensiones:

Tabla 3-3: Dimensiones de la matriz bidimensional

	Total filas	Total columnas
Adif	10954	748

Realizado por: Toapanta, Erika, 2022.

Para analizar la información de manera conjunta se procede a trabajar con la matriz bidimensional, sin embargo, para interpretar la información y crear un buen modelo de clustering se redujo la dimensionalidad de la información, teniendo en cuenta que la matriz tiene 10954 filas con 748 columnas. Se espera que en la reducción de la dimensionalidad se pueda representar el mayor porcentaje de la variabilidad de los datos.

3.2. Reducción de la dimensionalidad

Debido a que la matriz de información con la que se trabajará tiene 748 columnas, se escoge una técnica que nos ayude a reducir la dimensionalidad, y así trabajar de mejor manera con los grupos.

3.2.1. Componentes principales

Para esta técnica es necesario establecer el porcentaje de varianza que se desea explicar con las primeras componentes. En este caso se seleccionó el 90% lo que nos dio como resultado que para llegar a este número es necesario trabajar con un total de 31 componentes como se detalla en el gráfico:

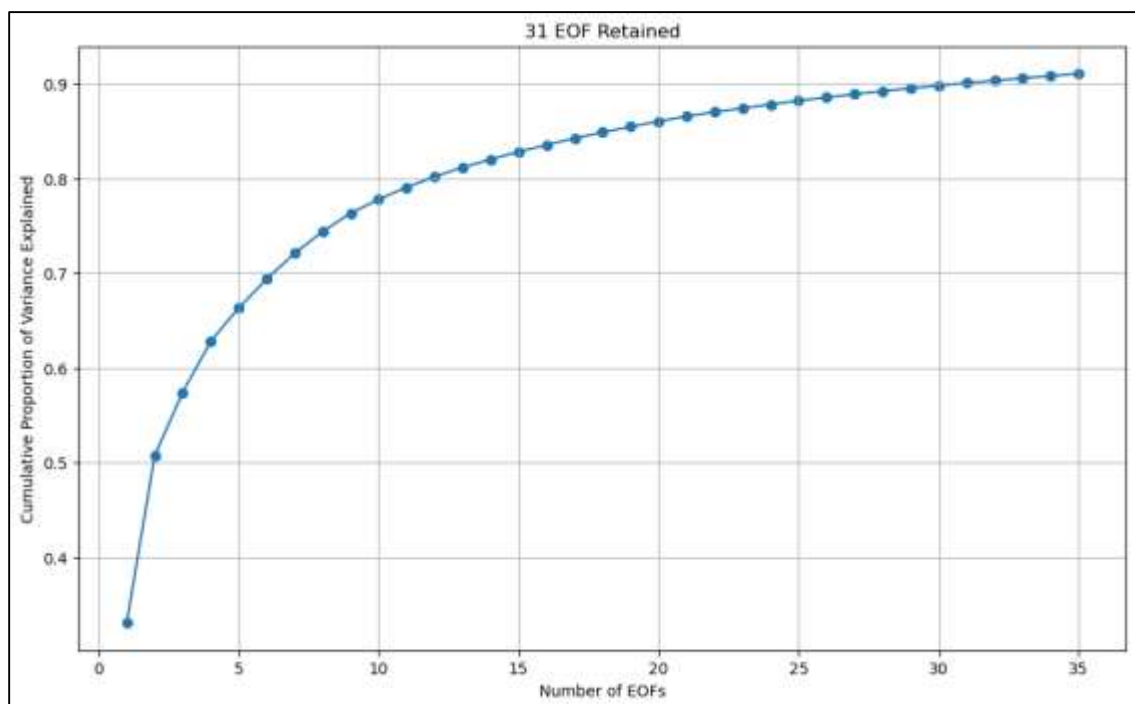


Gráfico 3-3. Elección del número de componentes principales.

Realizado por: Toapanta, Erika, 2022.

En la **Gráfica 4-3**, podemos observar que con las tres primeras componentes explicamos más del 50% de la variabilidad de los datos, desde la componente 5 el aporte de las componentes siguientes se vuelve insignificante.

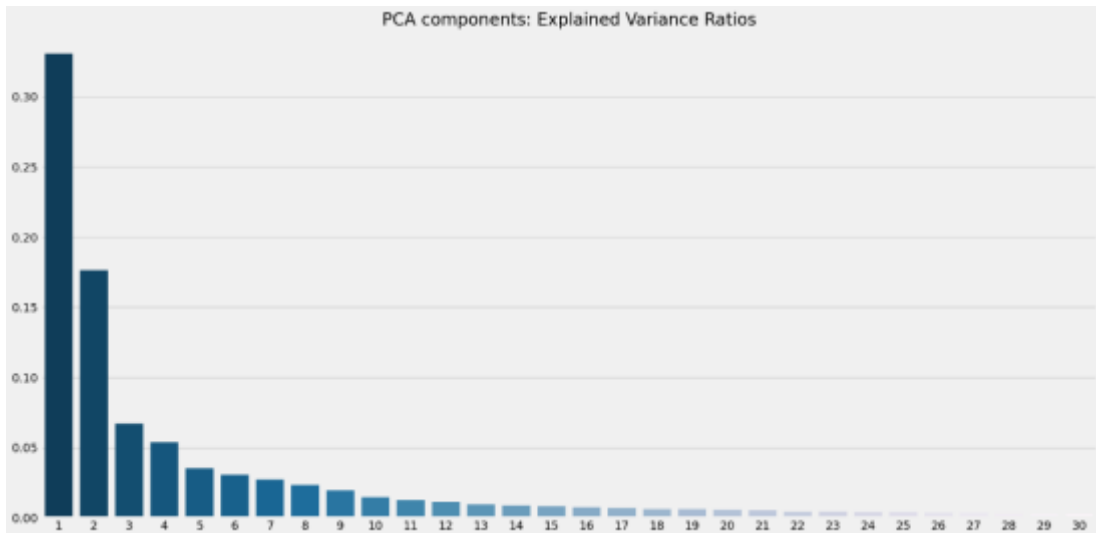


Gráfico 4-3. Porcentaje de varianza explicada.

Realizado por: Toapanta, Erika, 2022.

Ahora se representa las dos primeras componentes en un diagrama de dispersión el cual explica 50.74% de la variabilidad total.

3.2.2. *Diagrama de dispersión*

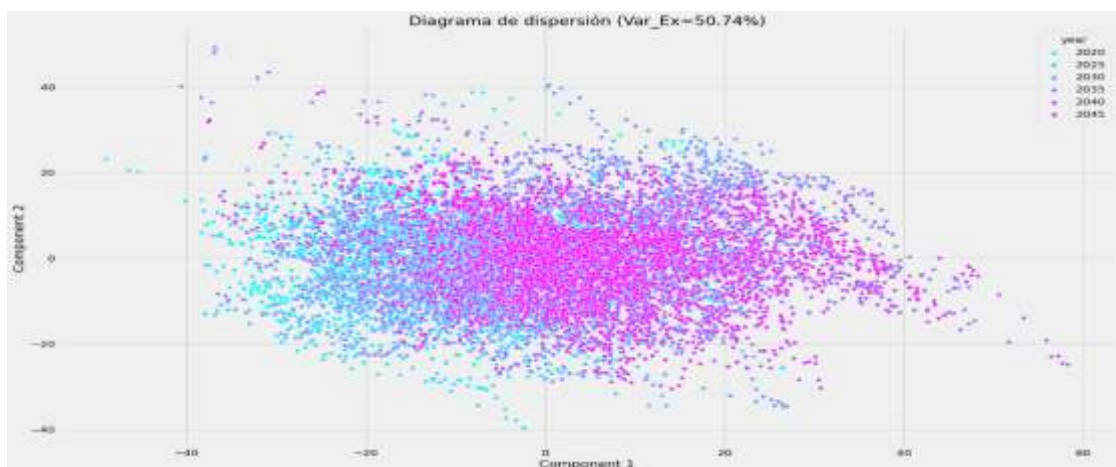


Gráfico 5-3. Diagrama de dispersión de las dos primeras componentes.

Realizado por: Toapanta, Erika, 2022.

En el **Gráfico 5-3**, se observa que la temperatura se comporta de manera uniforme en su mayoría salvo por algunas fechas que se alejan del comportamiento normal, el año no aporta mucho en estas dos dimensiones se ve que se sobreponen.

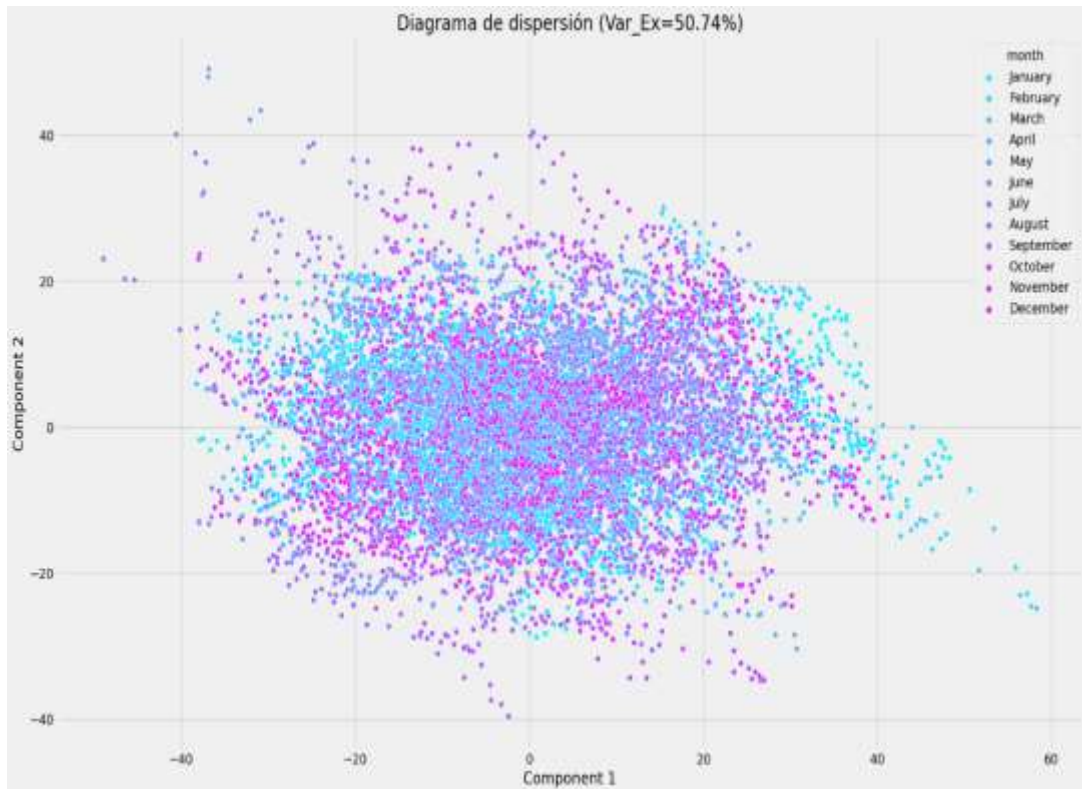


Gráfico 6-3. Diagrama de dispersión de las dos primeras componentes por mes.

Realizado por: Toapanta, Erika, 2022.

En el **Gráfico 6-3**, se evidencia que al etiquetarlo por mes, se vuelve aún más complejo visualizar algún tipo de comportamiento, únicamente se observa que los primeros meses del año se alejan un poco del resto.

3.2.3. *T-SNE*

Incrustación de vecinos estocásticos distribuidos en T, es otro de los algoritmos que se probarán para reducir la dimensionalidad de los datos:

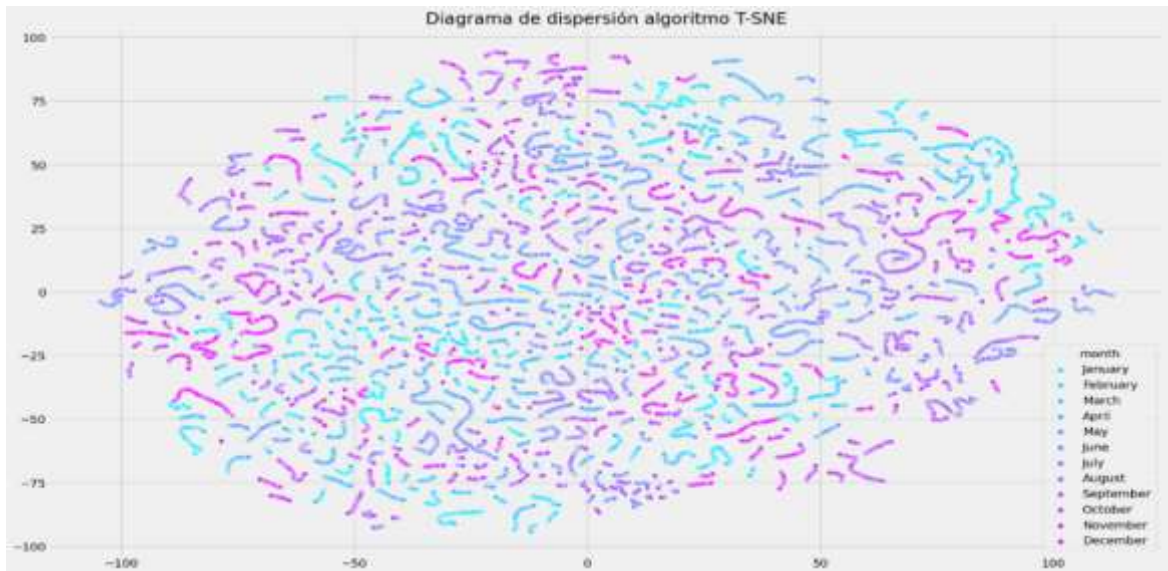


Gráfico 7-3. Incrustación de vecinos estocásticos distribuidos en t .

Realizado por: Toapanta, Erika, 2022.

En el **Gráfico 7-3**, es notable que el comportamiento de la temperatura tiene patrones no muy definidos, es decir no se puede observar similitudes entre grupos en periodo de meses.

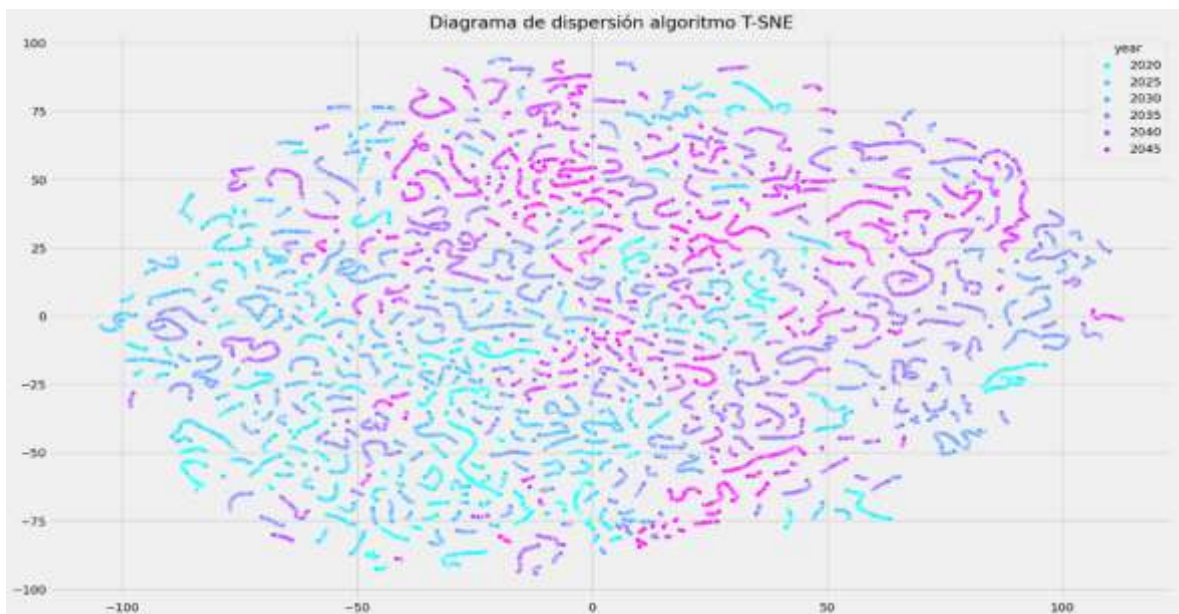


Gráfico 8-3. Incrustación de vecinos respecto a los años.

Realizado por: Toapanta, Erika, 2022.

En el **Gráfico 8-3**, utilizando esta técnica no se discrimina los grupos por sus similitudes, representarla por años no sería adecuada.

3.2.4. UMAP

Aproximación y proyección uniforme del colector, es otro de los algoritmos muy utilizados para reducir la dimensionalidad de los datos:

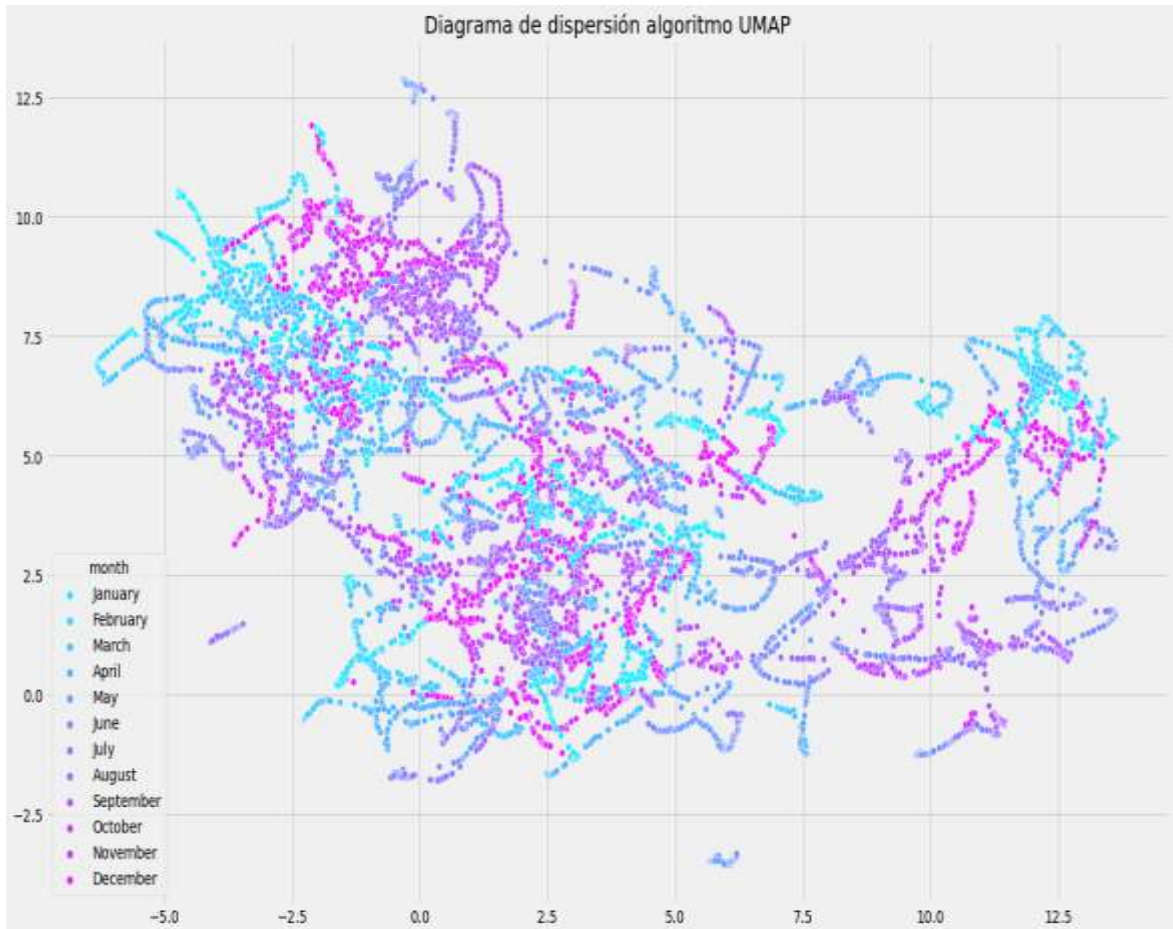


Gráfico 9-3. UMAP dimensionalidad

Realizado por: Toapanta, Erika, 2022.

En el **Gráfico 9-3**, mediante la aplicación del algoritmo UMAP se observa que este algoritmo separa los datos en tres grupos aproximadamente, posteriormente se procede a etiquetarlo por año.

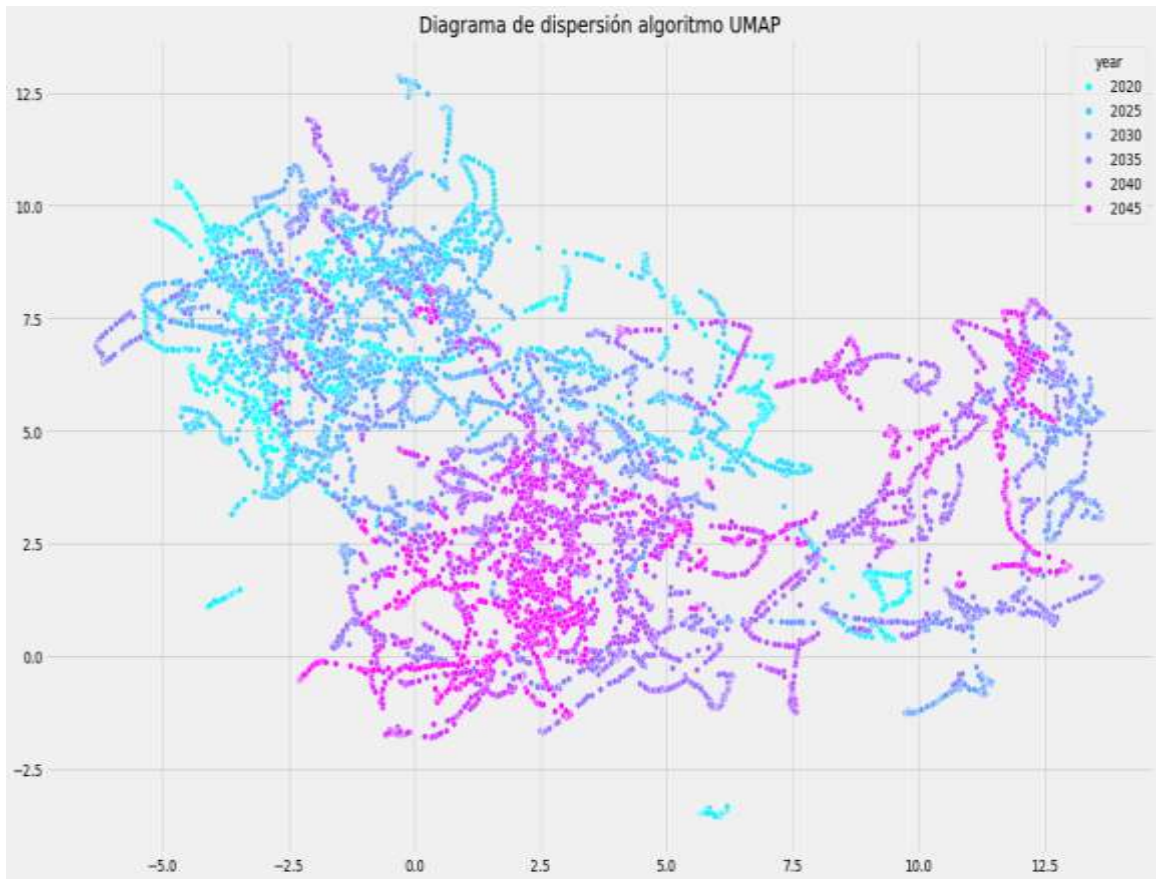


Gráfico 10-3. Umap clasificado por años

Realizado por: Toapanta, Erika, 2022.

En el **Gráfico 10-3**, se observa que efectivamente el resultado al etiquetar los puntos por año se separa los años, en términos de temperatura puede ser que posea años más cálidos o fríos que otros.

3.3. Algoritmos de clustering

A partir de la reducción de la dimensionalidad utilizando UMAP, se aplica el algoritmo k-means para la identificación de los patrones de comportamiento.

3.4.1. Método de Elbow o codo

El presente método utiliza los valores de la inercia donde se procede de la suma de las distancias al cuadrado de cada objeto del clúster a su centroide, para determinar el mejor número de grupos.

$$Inercia = \sum_{i=0}^N \|x_i - \mu\|^2$$



Gráfico 11-3. Gráfica de codo.

Realizado por: Toapanta, Erika, 2022.

En el **Gráfico 11-3**, de codo se observa que el numero óptimo de clústeres es 4, que puede estar clasificado por las regiones naturales del Ecuador (Región Interandina , Litoral, Amazónica, Insular).

3.3.1. Algoritmo *K Means* - UMAP con parámetros (métrica=" chebyshev")

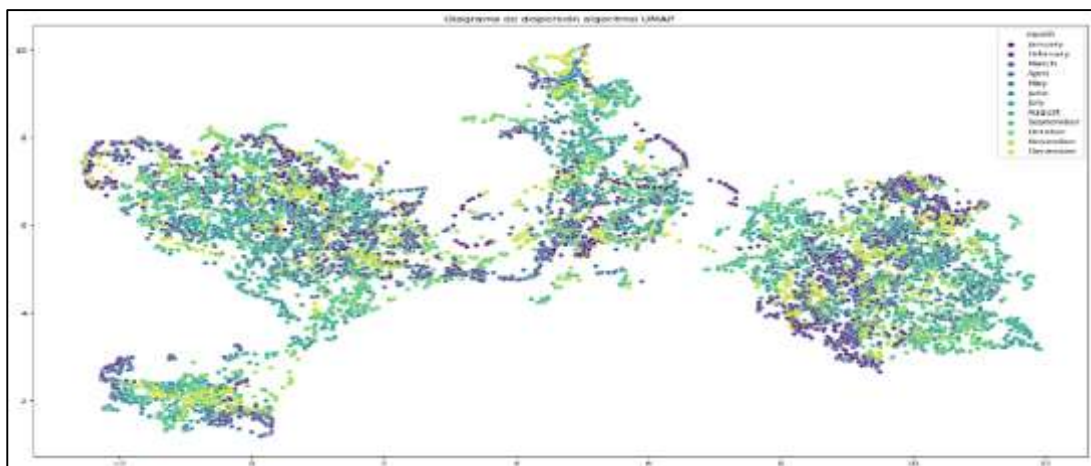


Gráfico 12-3. UMAP métrica chebyshev respecto meses.

Realizado por: Toapanta, Erika, 2022.

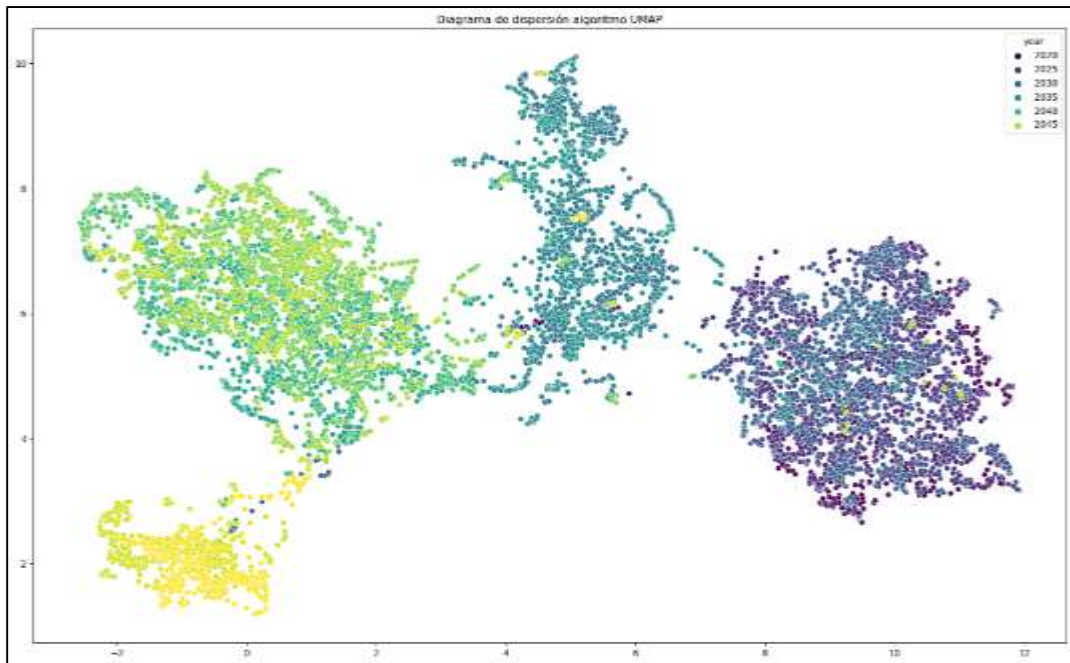


Gráfico 13-3. UMAP métrica chebyshev respecto años.

Realizado por: Toapanta, Erika, 2022.

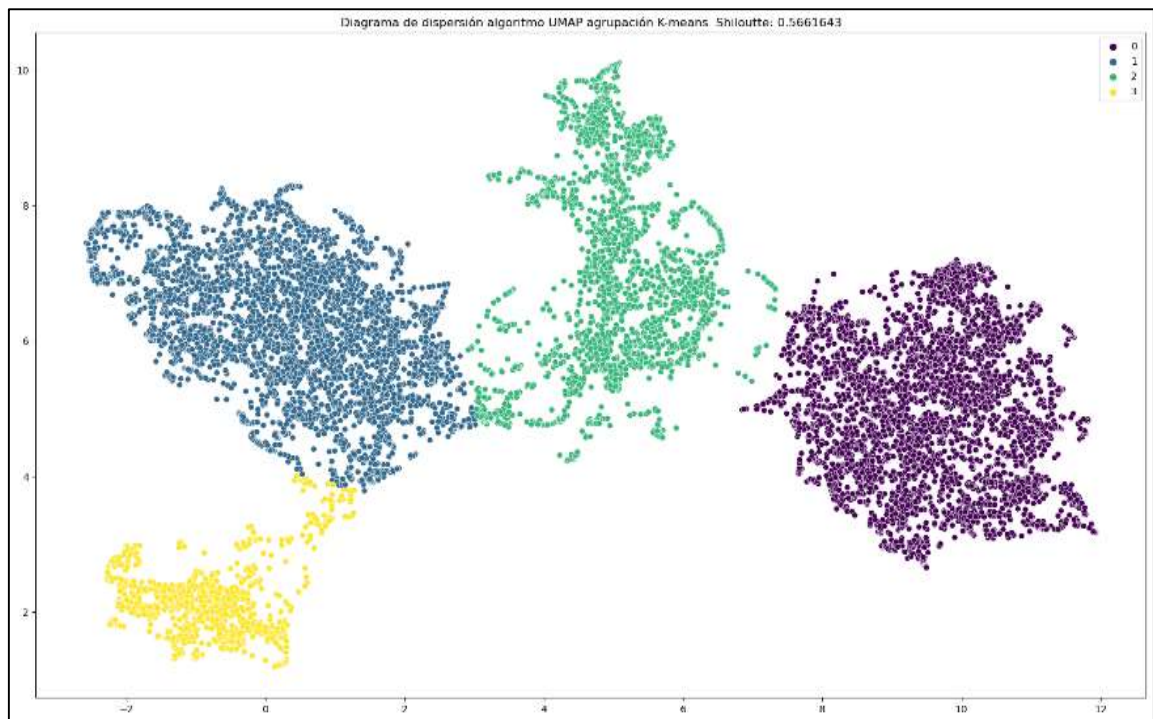


Gráfico 14-3. UMAP métrica chebyshev agrupación K-means

Realizado por: Toapanta, Erika, 2022.

Finalmente, en el **Gráfico 14-3**, se obtuvo un conjunto de 4 grupos bien definidos mediante la técnica UMAP con la métrica de Chebyshev, los grupos forman los siguientes rangos de años:

3.3.2. Agrupación de datos según el rango de años

Tabla 4-3: Rango de años para los grupos.

Rango de años	Grupo	Cantidad de observaciones
2020-2030	0	4016
2031-2035	1	3653
2036-2045	2	1826
2046-2049	3	1459

Realizado por: Toapanta, Erika, 2022.

3.3.3. Matriz de confusión

Tabla 5-3: Matriz de confusión

		Valores Reales			
		2020-2030	2031-2035	2036-2045	2046-2049
Valores predichos	2020-2030	3692	9	303	12
	2031-2035	21	3367	224	41
	2036-2045	46	23	1751	6
	2046-2049	12	342	35	1070
Accuracy		0.901953			

Realizado por: Toapanta, Erika, 2022.

En la **Tabla 5-3**, es notable que por la matriz de confusión y la exactitud calculada, tenemos un buen modelo, mediante la elección de la técnica UMAP con el Algoritmo K-means , mediante las distancias de Chebyshev que establece un 90% de precisión que indica que las observaciones se encuentran bien clasificadas en cada grupo, además se indica en este resultado que en ciertos años ocurre un desfase del clima, es decir que los rangos de años que se eligieron se comportan de manera diferente unos de otros según los 4 grupos .

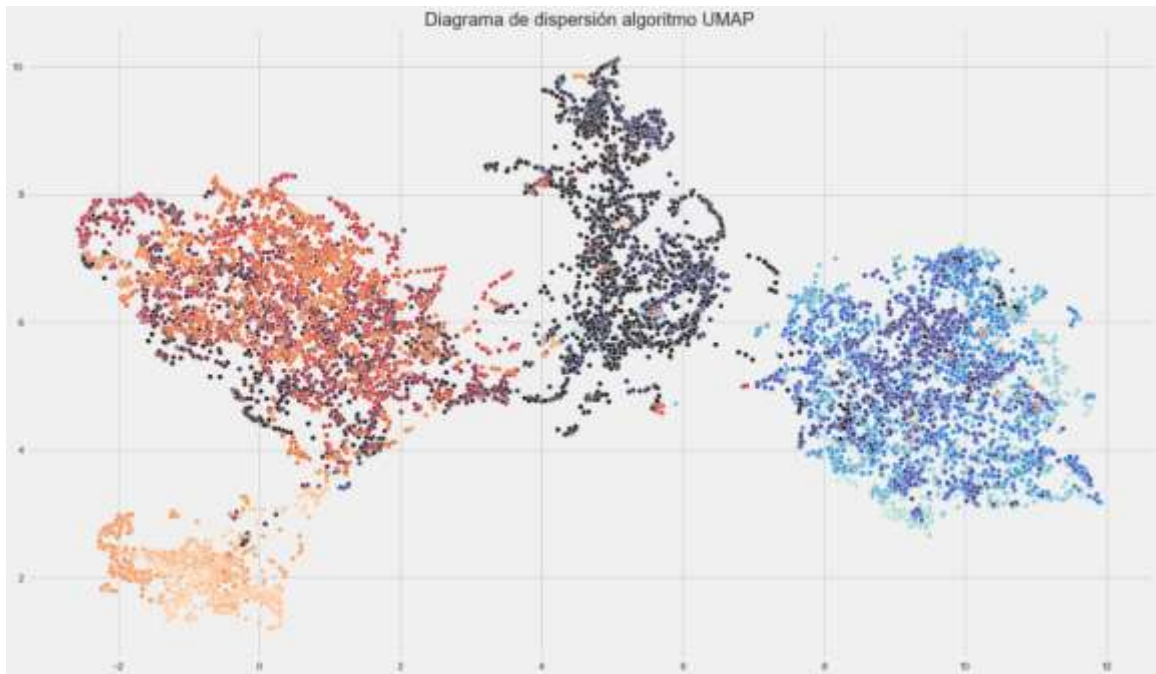


Gráfico 15-3. Diagrama de dispersión UMAP

Realizado por: Toapanta, Erika, 2022.

Visualizamos que en los días se encuentran bien clasificados en los 4 grupos, este gráfico nos muestra que según avanza el tiempo, existe una variación en la temperatura y por lo tanto esto se va estar reflejado en el comportamiento de los patrones, la causa puede ser desde un fenómeno climático hasta cierta frecuencia de días en los que ocurre este patrón como puede ser: temperaturas bajas, temperaturas altas, temperaturas normales y temperaturas muy altas.

3.3.4. Representación de resultados por días del algoritmo K-means métrica chebyshev.

Tabla 6-3: Rango de años para los grupos.

Rango de años	Grupo	Cantidad de observaciones predichas	Frecuencia Relativa	Patrones
2020-2030	0	3771	34.43%	Temperaturas normales
2031-2035	1	3741	34.15%	Temperaturas altas
2036-2045	2	2313	21.12%	Temperaturas bajas
2046-2049	3	1129	10.31%	Temperaturas muy altas

Realizado por: Toapanta, Erika, 2022.

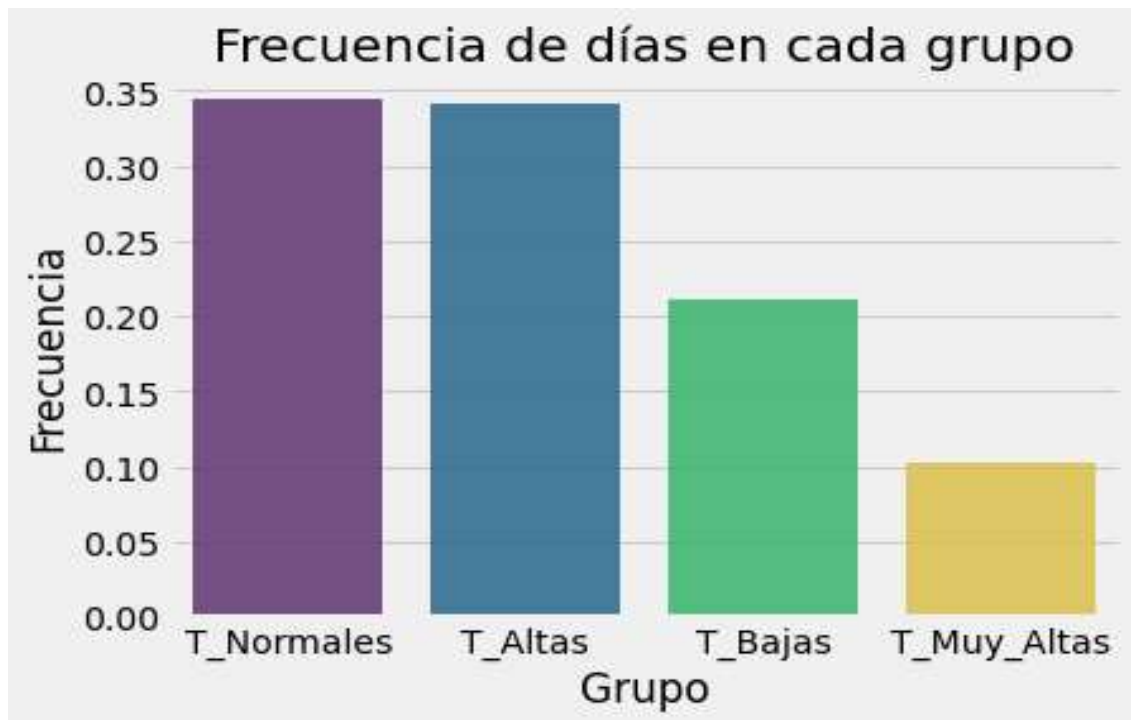


Gráfico 16-3: Diagrama de frecuencias temperaturas

Realizado por: Toapanta, Erika, 2022.

- Las temperaturas **“normales”** correspondiente al grupo 0, presentan una mayor ocurrencia del 34,43%, desde el año 2020-2030, donde se evidencia que no existirán cambios significativos en la variabilidad del comportamiento de la temperatura.
- Las temperaturas **“altas”** respecto al grupo 1, presentan una mayor ocurrencia del 34,15%, en el periodo 2031-2035, donde se estima periodos de sequías.
- Las temperaturas **“bajas”** corresponden al grupo número 2, representado por el 21,12%, en el periodo 2036-2045 donde se anuncian lluvias frecuentes.
- Las temperaturas **“muy altas”** corresponden al grupo número 3, presentado con menor frecuencia del 10,31%, respecto al año 2046-2049 evidenciando aún más el cambio climático generado por el calentamiento global siendo motivo de a lo que sucederá en el futuro.

3.4. Mapas obtenidos

Se realiza un mapa con la temperatura promedio de cada año desde el 2020 al 2050, el cual nos permite observar el comportamiento de esta variable en la naturaleza.

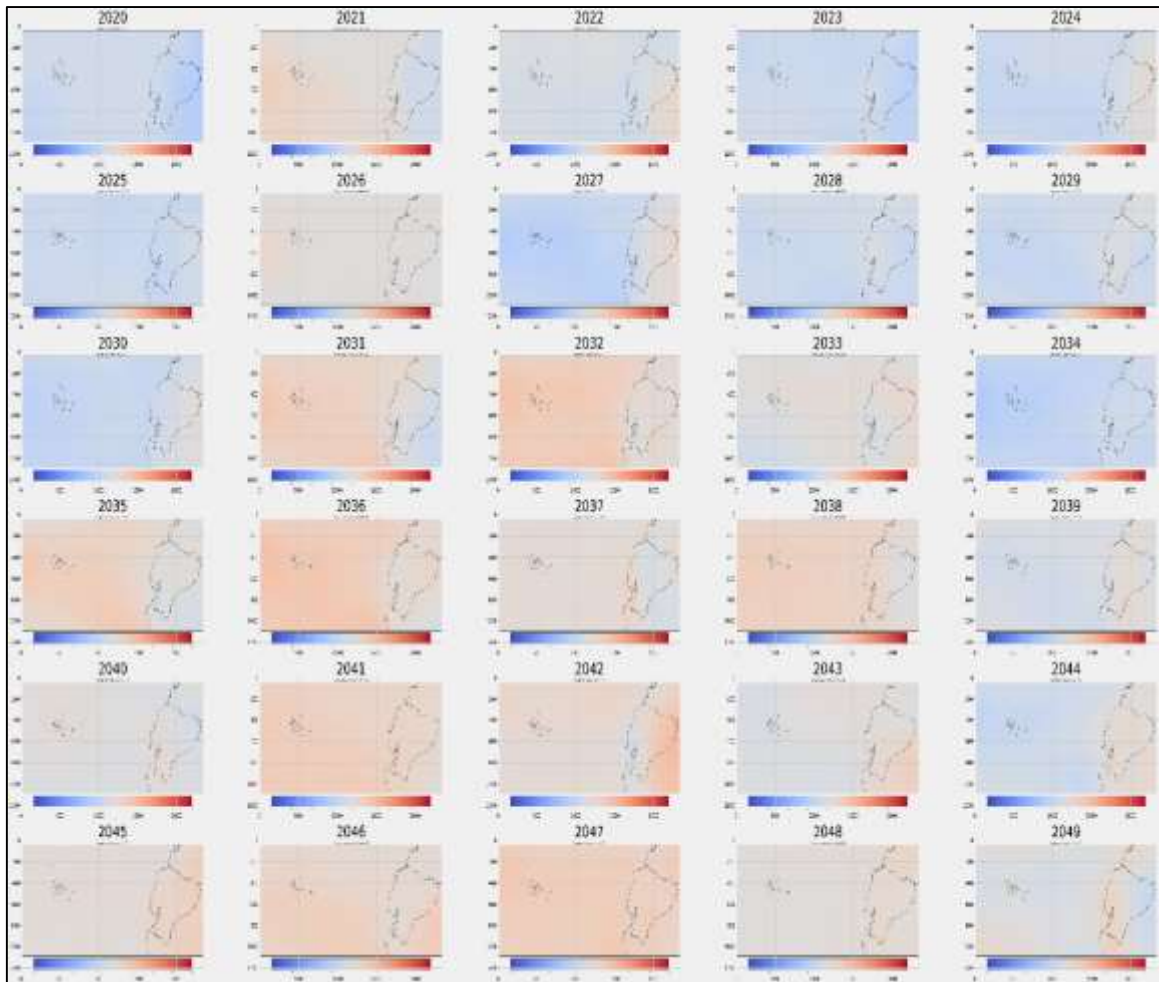


Gráfico 17-3. Mapas generados con la temperatura promedio para cada año.

Realizado por: Toapanta, Erika, 2022.

Se observa en el **Gráfico 17-3**, respecto a los años 2020-2050, el comportamiento de la temperatura varía en ciertos años, uno de los posibles motivos es debido a la ubicación en la línea ecuatorial y su influencia de la humedad procedente de la Amazonía siendo nuestro país vulnerable a las alteraciones ocasionadas por el calentamiento global, así mismo identificamos el cambio que se genera en algunas regiones donde son mucho más cálidos que otros.

Gráfica de calor

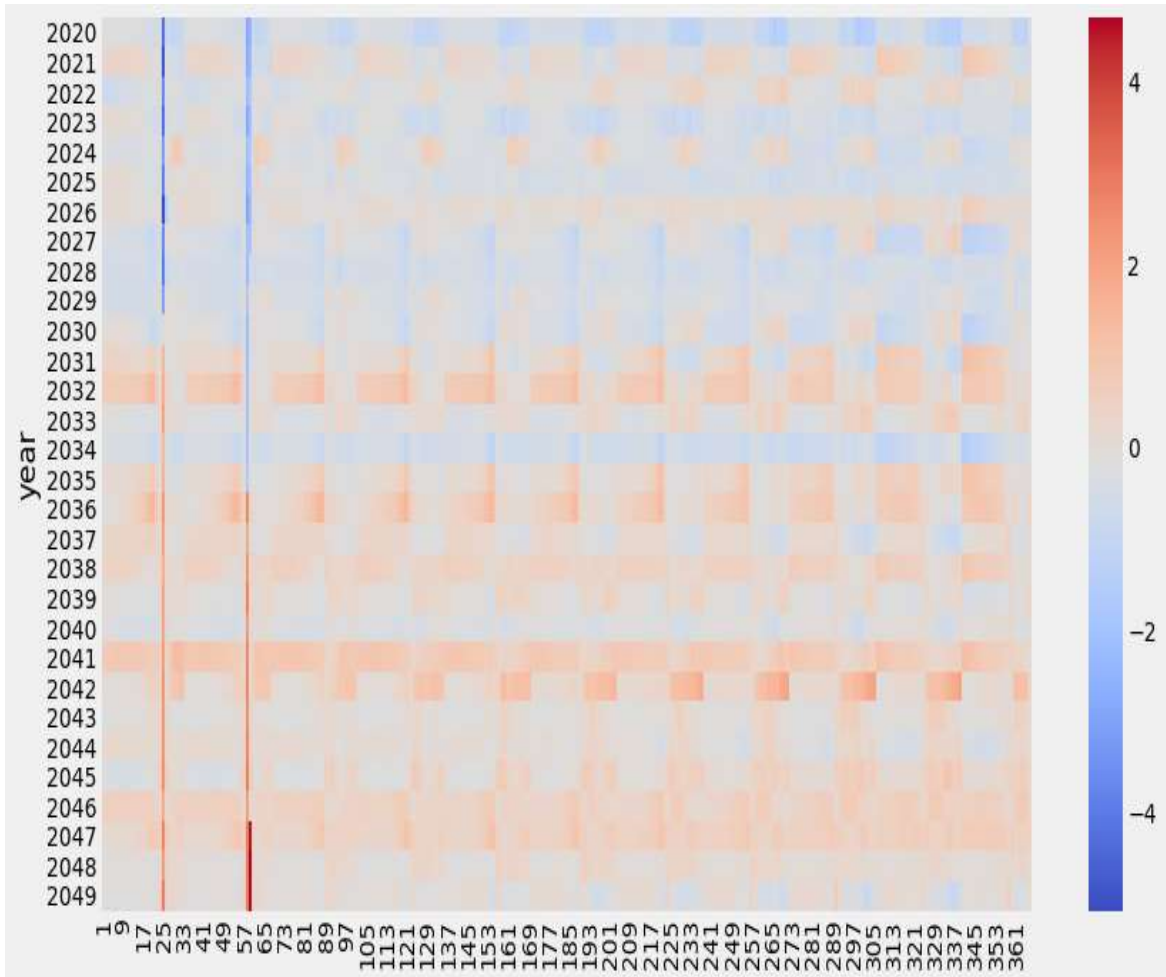


Gráfico 18-3. Gráfico de Calor representado diariamente respecto a los años

Realizado por: Toapanta, Erika, 2022.

En el gráfico de calor representado, se puede observar que entre los años 2047 al 2050 preexistirá un incremento en la temperatura. Cabe destacar que en el mes de febrero se presentará un comportamiento abrupto.

3.4.1. *Temperaturas Altas (2031-2035)*

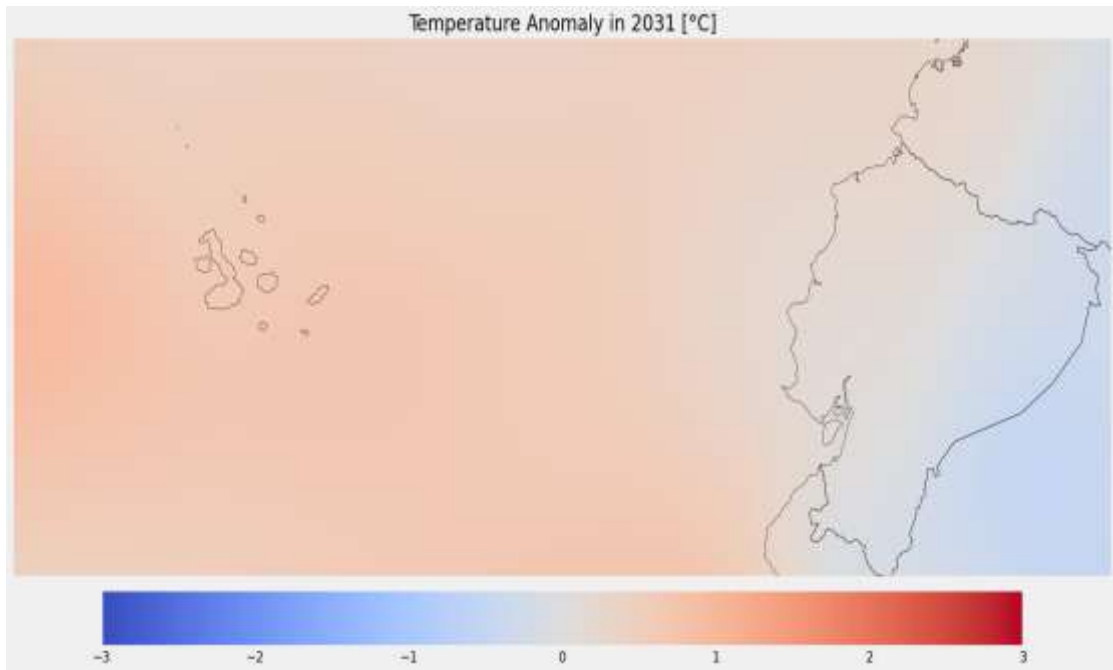


Gráfico 19-3. Mapa generado con la temperatura promedio correspondiente al 2031.

Realizado por: Toapanta, Erika, 2022.

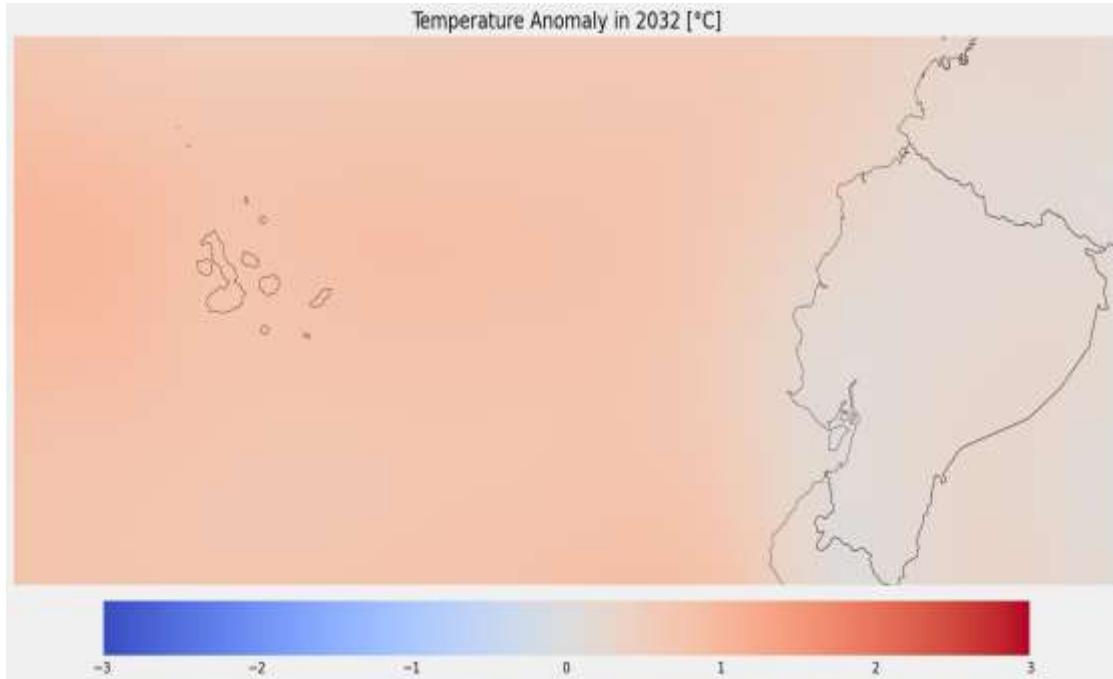


Gráfico 20-3. Mapa generado con la temperatura promedio correspondiente al 2032.

Realizado por: Toapanta, Erika, 2022.

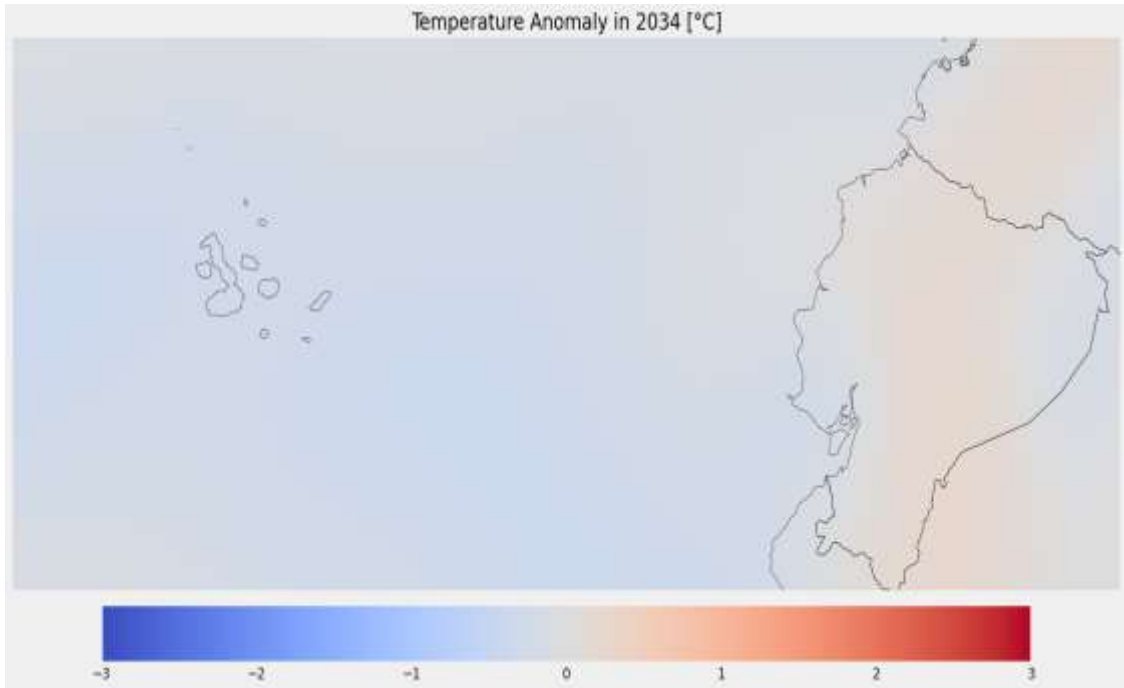


Gráfico 21-3. Mapa generado con la temperatura promedio correspondiente al 2034.

Realizado por: Toapanta, Erika, 2022.

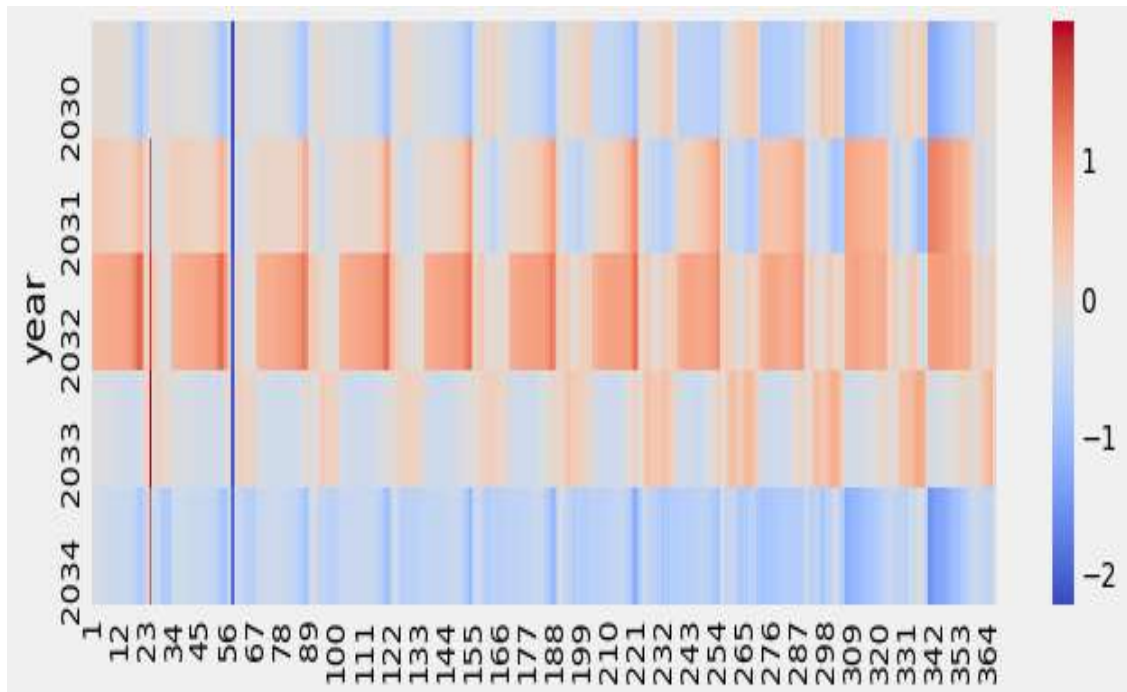


Gráfico 22-3. Gráfico de Calor representado en los últimos 4 años

Realizado por: Toapanta, Erika, 2022.

El cambio climático se evidencia en el periodo 2031-2032, la temperatura aumenta abruptamente, lo que generaría sequías prolongadas durante todo el, cuantiosas pérdidas: económicas y recursos naturales en todo el Ecuador. En el año 2034 en el mes de enero desde el día 23 aproximadamente hasta el mes de febrero se prevé que la temperatura se incrementará por ende es necesario

compartir estos resultados al sector agropecuario y así puedan mitigar estos efectos tomando decisiones oportunas.

3.4.2. *Temperaturas Bajas (2036-2045)*

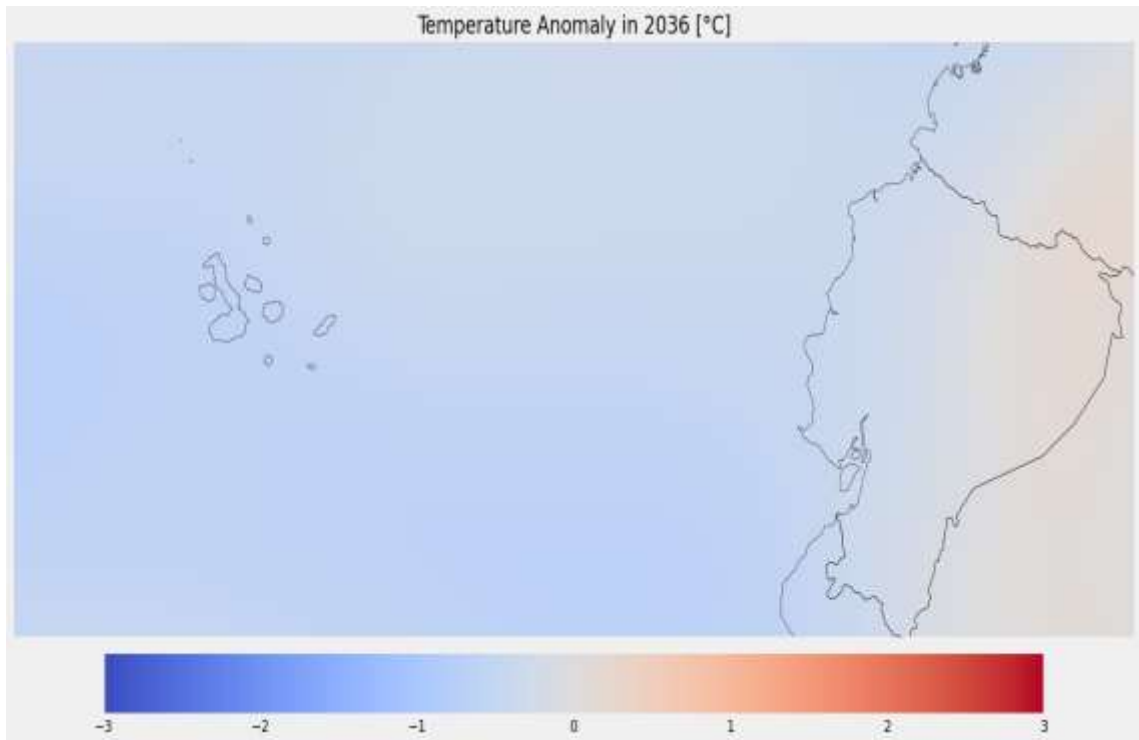


Gráfico 23-3. Mapa generado con la temperatura promedio correspondiente al 2036.

Realizado por: Toapanta, Erika, 2022.

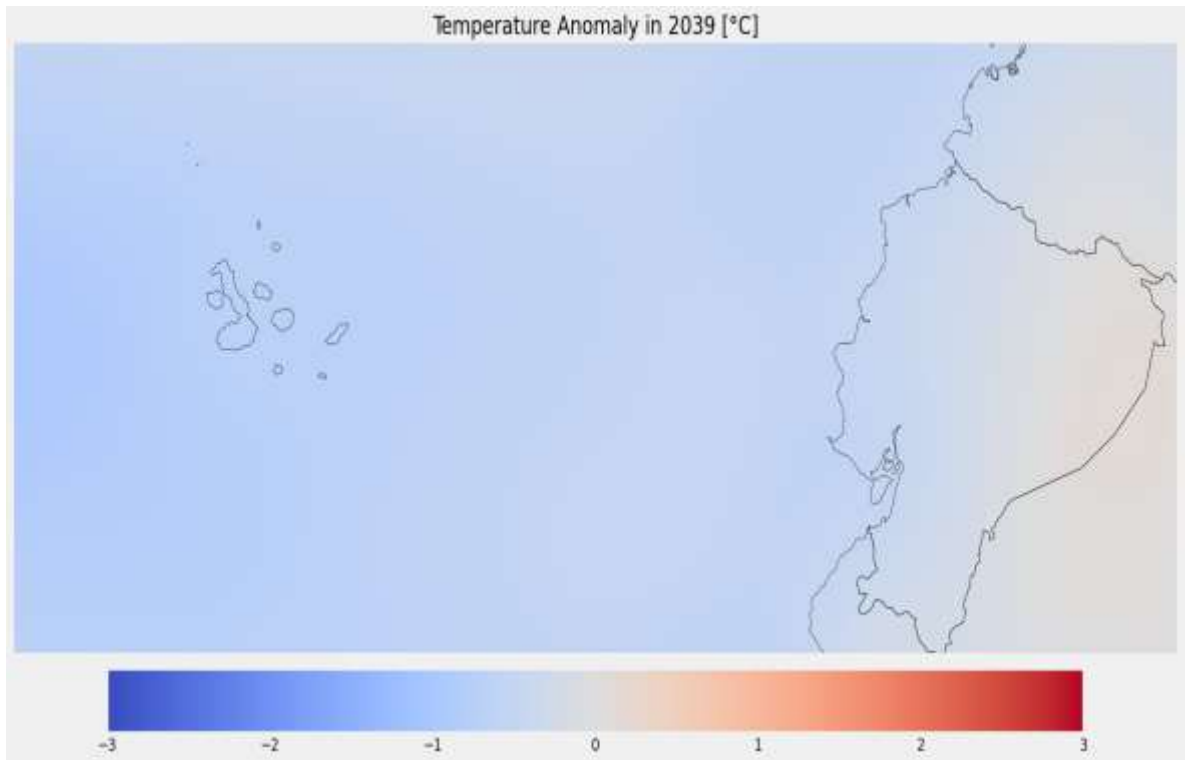


Gráfico 24-3. Mapa generado con la temperatura promedio correspondiente al 2039.

Realizado por: Toapanta, Erika, 2022.

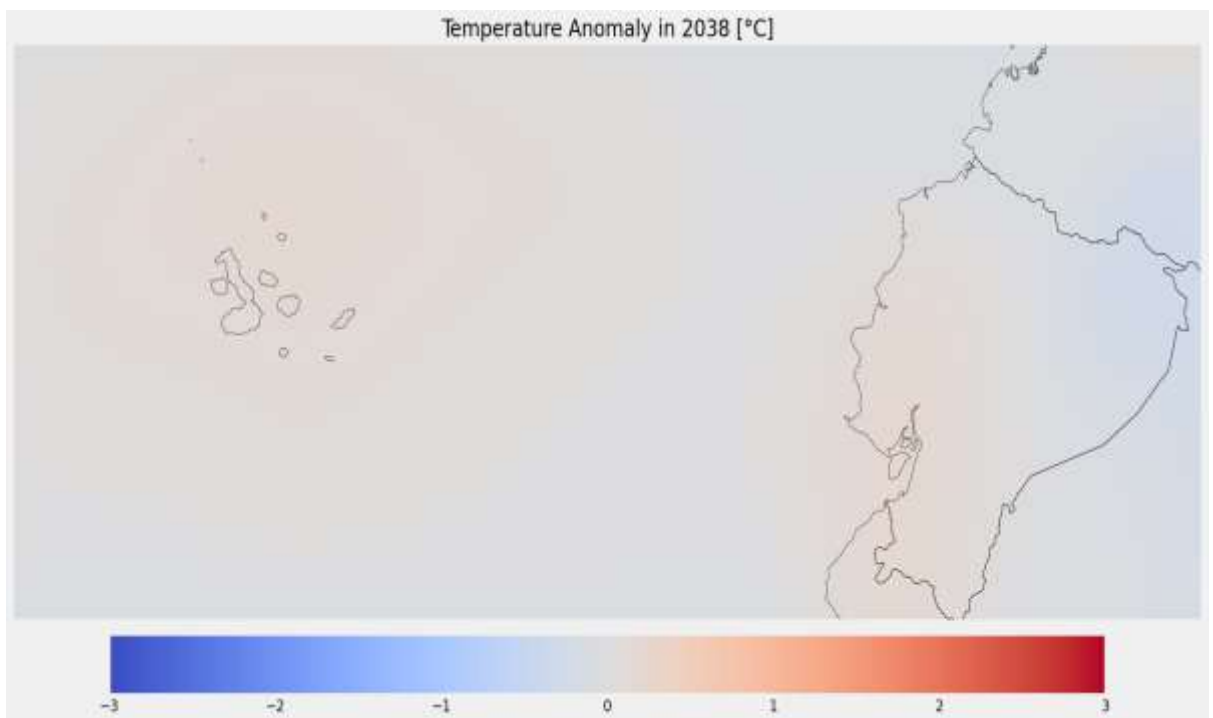


Gráfico 25-3. Mapa generado con la temperatura promedio correspondiente al 2038.

Realizado por: Toapanta, Erika, 2021.

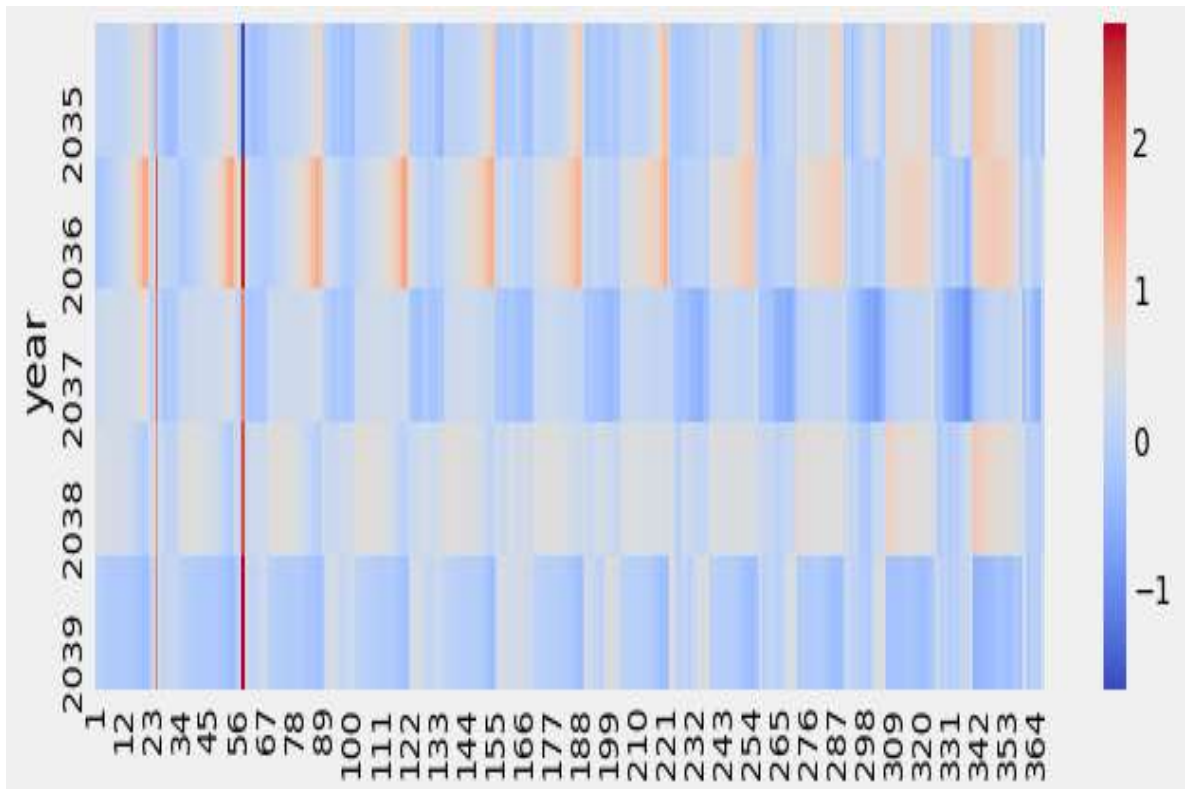


Gráfico 26-3. Grafico de Calor representado en los últimos 4 años

Realizado por: Toapanta, Erika, 2022.

En el **Gráfico 26-3**, se evidencia, en el año 2036 el tiempo climático varía en los meses de enero, febrero y marzo, con un incremento en la temperatura en la Región de la Amazonía.

Sin embargo, desde el año 2037 al 2045 la temperatura para el resto del Ecuador es baja, la presente investigación debería ser compartida al Ministerio de Salud para la toma de decisiones para alertar a la población y tomar medidas necesarias.

3.5. Algoritmos que fueron probados para su selección

3.5.1. Algoritmo K-means - UMAP con parámetros (métrica="euclidean")

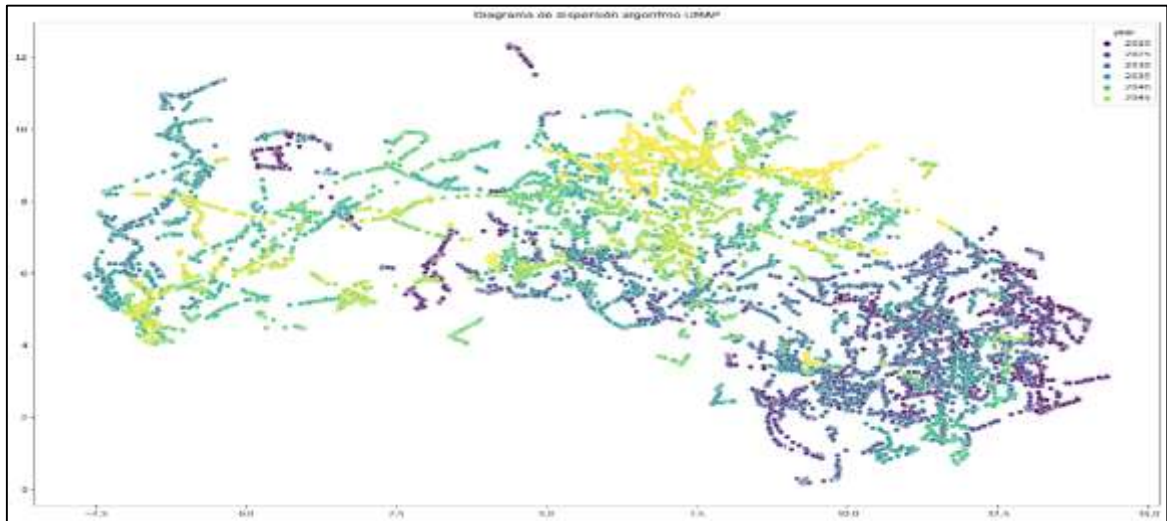


Gráfico 27-3. UMAP métrica euclídea respecto años

Realizado por: Toapanta, Erika, 2022.

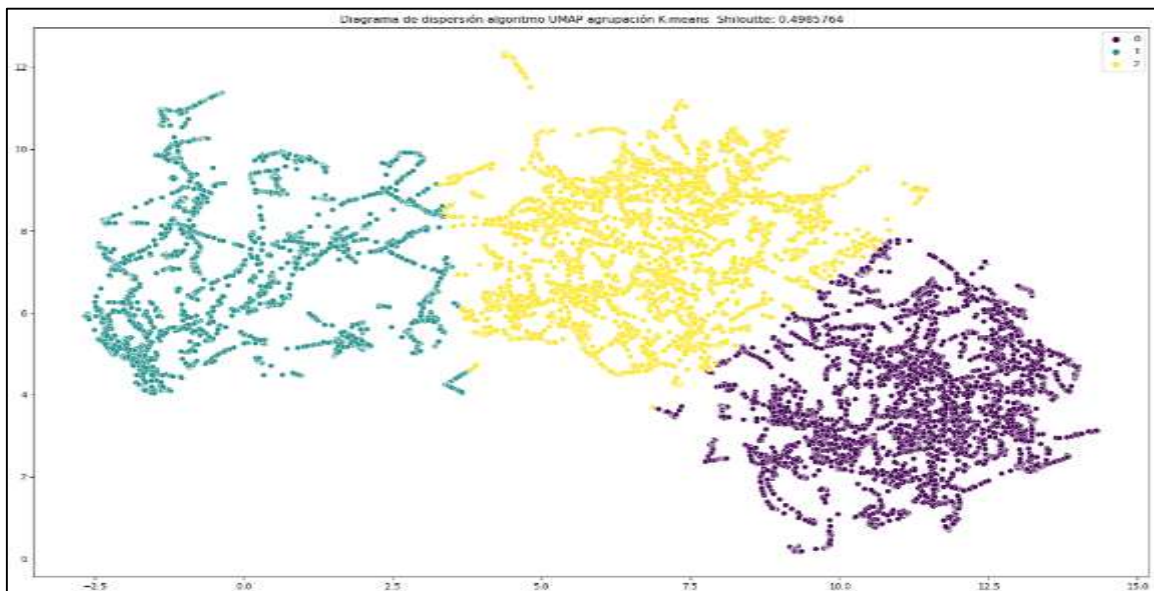


Gráfico 28-3. UMAP métrica euclídea agrupación K-means

Realizado por: Toapanta, Erika, 2022.

Es claro el resultado, utilizando UMAP con la métrica euclidiana ya que separa la información en 3 grupos, finalmente se validan los algoritmos para tener una perspectiva clara.

3.5.2. Algoritmo DBSCAN - UMAP con parámetros (metric="euclidean")

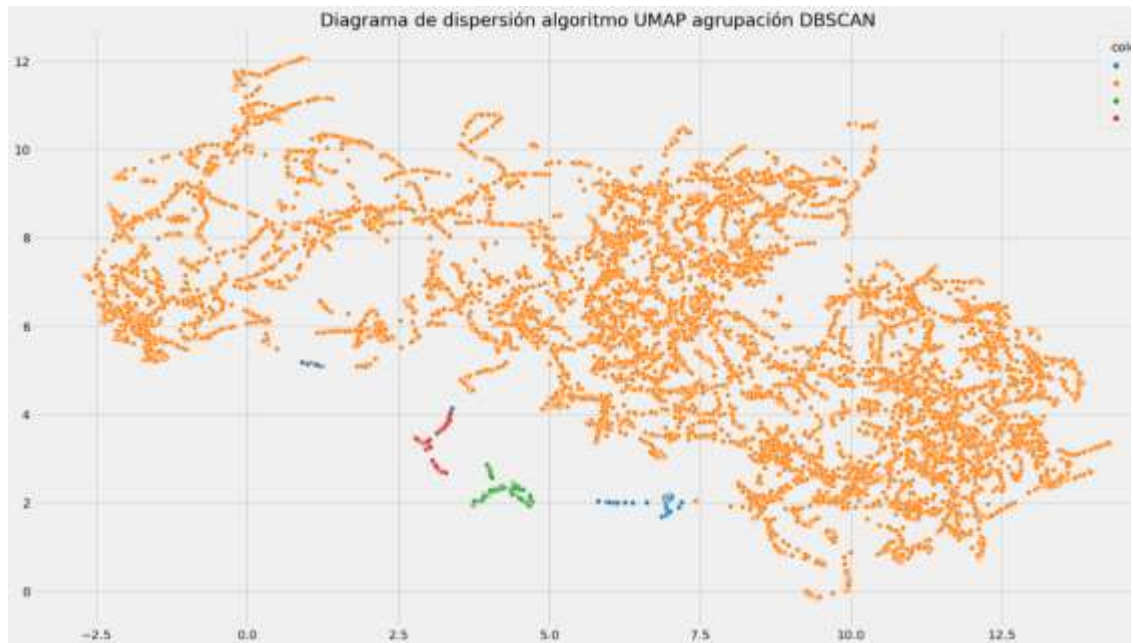


Gráfico 29-3. UMAP métrica euclídea agrupación DBSCAN.

Realizado por: Toapanta, Erika, 2022.

3.5.3. Algoritmo DBSCAN - UMAP con parámetros (métrica="chebyshev")

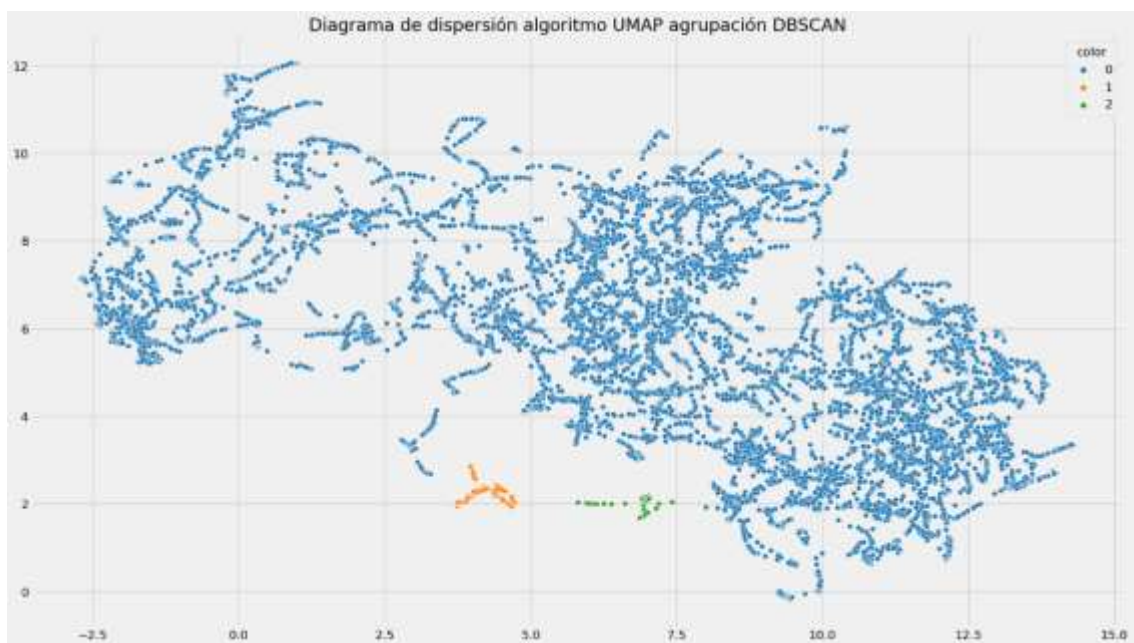


Gráfico 30-3. UMAP métrica euclídea agrupación DBSCAN.

Realizado por: Toapanta, Erika, 2022.

Se observa en las gráficas bajo el algoritmo DBSCAN con la métrica chebyshev, el comportamiento de los datos no poseen una buena agrupación lo que nos alejaría de tener una buena interpretación .

3.5.4. Algoritmo AGGLOMERATIVE CLUSTERING - UMAP con parámetros (métrica="euclidean")

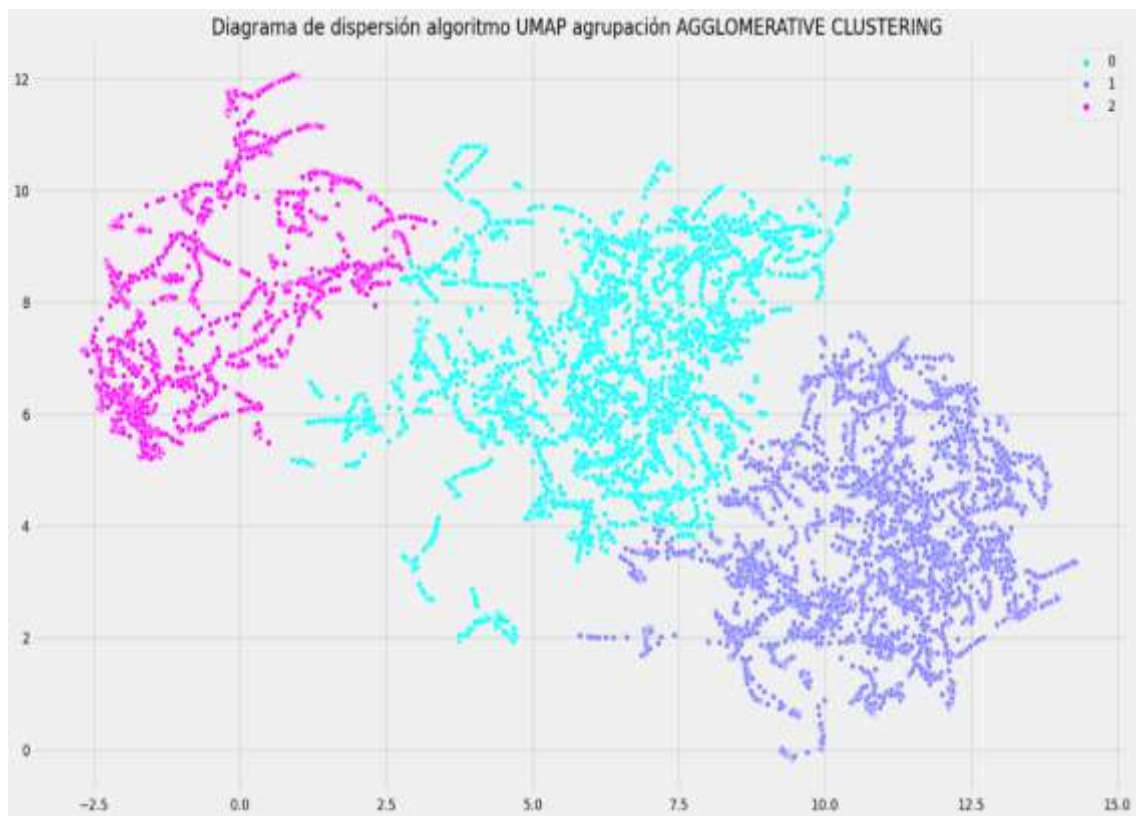


Gráfico 31-3. UMAP métrica euclídea agrupación Agglomerative Clustering.

Realizado por: Toapanta, Erika, 2022.

3.5.5. Algoritmo *AGGLOMERATIVE CLUSTERING* - *UMAP* con parámetros (*métrica="chebyshev"*)

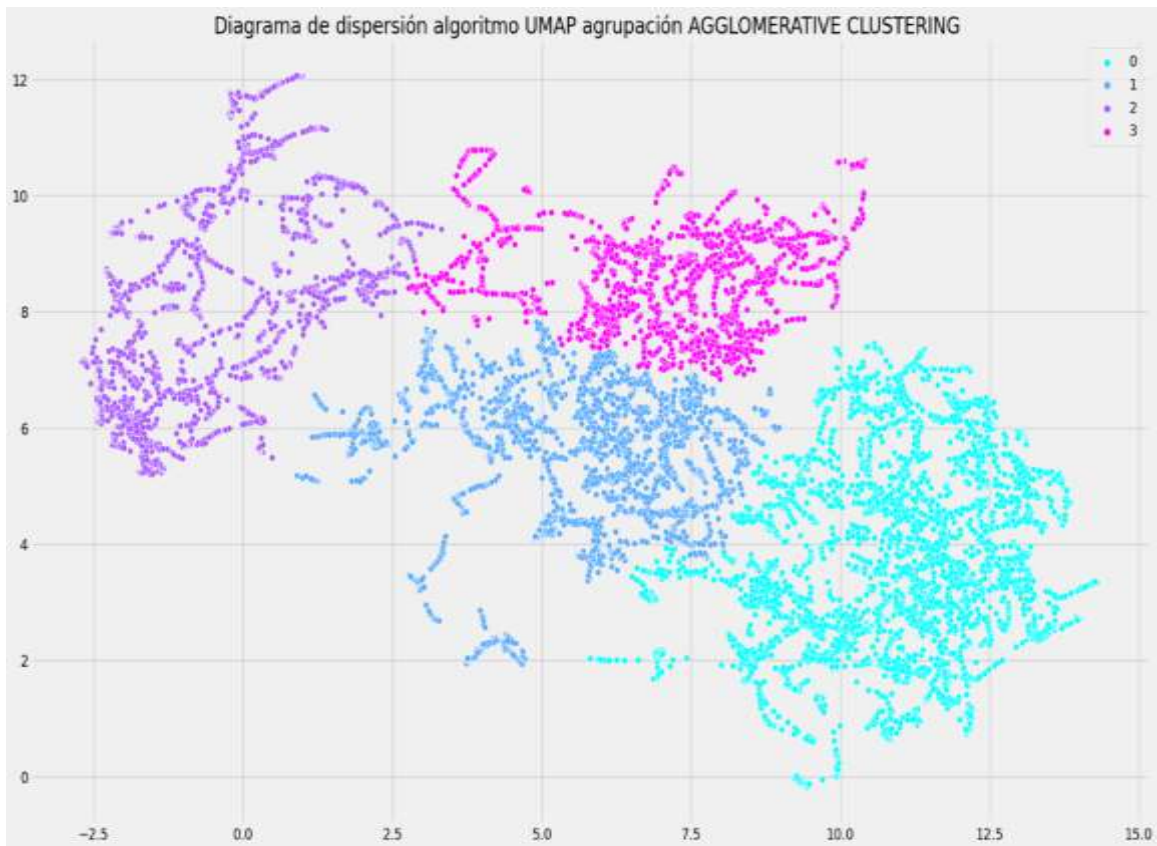


Gráfico 32-3. UMAP métrica euclídea agrupación Agglomerative Clustering.

Realizado por: Toapanta, Erika, 2022.

Observamos que mediante el algoritmo Agglomerative clustering los datos se encuentran bien distribuidos en los 4 grupos, sin embargo, se debe validar el modelo para tener un resultado confiable.

3.5.6. Validación del modelo

Mediante el gráfico de código se conoció a priori que el número de grupos a trabajar es 4, para contrastar y validar el modelo se realiza el cálculo del índice de Silhouette y obtenemos como resultado lo siguiente:

Tabla 7-3: Índices de *Silhouette*

Características UMAP	Algoritmo	Grupos	Silhouette
UMAP metric="euclidean"	k-means	3	0.4985
UMAP metric="chebyshev"	k-means	4	0.7217
UMAP metric="euclidean"	DBSCAN	2	0.1814
UMAP metric="chebyshev"	DBSCAN	4	0.5373
UMAP metric="euclidean"	Agglomerative Clustering	3	0.4652
UMAP metric="chebyshev"	Agglomerative Clustering	4	0.56164

Realizado por: Toapanta, Erika, 2022.

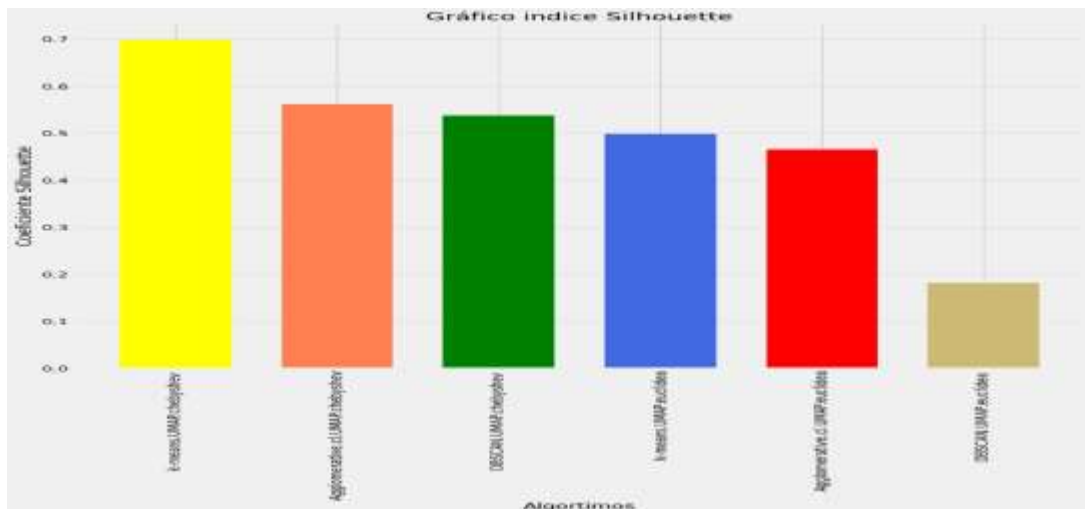


Gráfico 33-3. Diagrama de barras representando a los algoritmos y métricas.

Realizado por: Toapanta, Erika, 2022.

Se aprecia en el **Gráfico 33-3**, los valores de los índices de Silhouette que se hallaron en base a los diferentes algoritmos de agrupación y además de las distintas características del algoritmo UMAP, la característica que mejor da resultados es al cambiar la medida de distancia de la euclídea a la de Chebyshev, con una efectividad del 72%, los demás funcionan casi igual de bien que k-means pero el algoritmo k-means se recomienda usar cuando los grupos tienen una forma más o menos circular, por lo que se selecciona este algoritmo para realizar el cálculo de la exactitud del mismo frente a la agrupación por años.

3.5.7. Validación del modelo

Tabla 8-3: Temperaturas registradas en el Ecuador mes de mayo del año 2021.

Región	Temperatura superficial máxima del aire	Temperatura superficial mínima del aire
Región Litoral	Bajo la Normal	Baja la Normal
Región Interandina	Entre bajo la Normal y Normal	Sobre la normal a excepción de estaciones puntuales.
Región Amazónica	Normal	Bajo la normal, en la estribación de la cordillera, normal en la llanura Amazónica.
Región Insular	Normal	Bajo la normal

Fuente: INOCAR, 2021.

Tabla 9-3: Temperaturas registradas en el Ecuador mes de junio del año 2021.

Región	Temperatura superficial máxima del aire	Temperatura superficial mínima del aire
Región Litoral	Normal	Normal
Región Interandina	Entre bajo la Normal y Normal	Sobre la normal a excepción de estaciones puntuales.
Región Amazónica	Normal	Bajo la normal, en la estribación de la cordillera, normal en la llanura Amazónica.
Región Insular	Entre bajo la Normal	Sobre la normal

Fuente: INOCAR, 2021.

Tabla 10-3: Temperaturas registradas en el Ecuador mes de julio del año 2021.

Región	Temperatura superficial máxima del aire	Temperatura superficial mínima del aire
Región Litoral	Entre bajo la Normal y Normal	Sobre la normal a excepción de estaciones puntuales.
Región Interandina	Entre bajo la Normal y Normal	Sobre la normal a excepción de estaciones puntuales.
Región Amazónica	Normal	Bajo la normal, en la estribación de la cordillera, normal en la llanura Amazónica.
Región Insular	Normal	Bajo la normal

Fuente: INOCAR, 2021.

Tabla 11-3: Temperaturas registradas en el Ecuador mes de agosto del año 2021.

Región	Temperatura superficial máxima del aire	Temperatura superficial mínima del aire
Región Litoral	Bajo la Normal	Baja la Normal
Región Interandina	Entre bajo la Normal y Normal	Sobre la normal a excepción de estaciones puntuales.
Región Amazónica	Sobre la normal	Sobre la normal
Región Insular	Normal	Bajo la normal

Fuente: INOCAR, 2021.

La información del tiempo climático y su variación respecto al año 2021 de la variable temperatura se lo obtuvo de la página oficial del INNOCAR (www.inocar.mil.ec). En efecto la predicción del tiempo de la temperatura son análogos a los pronósticos de patrones de comportamiento mediante la utilización de la técnica Machine Learning bajo el algoritmo KMEANS.

CONCLUSIONES

- Se realizó una exploración de datos donde se describió la matriz de información, su formato, estructura multidimensional constituida por etiquetas de ubicación espacial y tiempo, donde se efectuó la apilación de las matrices en una sola matriz bidimensional.
- Se efectuó diversas técnicas de reducción de dimensionalidad una de ellas componentes principales donde las 5 primeras componentes manifiestan el 90% de la variabilidad explicada, además otras técnicas machine learning como el PCA, TSNE que no fueron lo suficientemente adecuados para su elección, finalmente UMAP siendo el mejor en medir las distancias entre las observaciones.
- Mediante la aplicación de los diferentes algoritmos de clustering “K-means, DBSCAN, Aglomerative Clustering” se trabajó con el que nos brindó mayor eficiencia en la agrupación de grupos optando por K-means ya que se pudo visualizar los patrones de comportamiento de la variable temperatura a lo largo del tiempo.
- Los métodos utilizados “**k-means métrica Chebyshev**” brindaron un mejor modelo de clustering compuesto por 4 grupos lo que ha generado en este estudio una precisión del 70% lo que indica que son buenos estimadores para describir el comportamiento climático en el Ecuador.
- Se validó los resultados del algoritmo con el Informe climático presentado por el INHAMI del año 2021 donde se ha evidenciado una temperatura máxima y mínima con un rango NORMAL, donde efectivamente siguen el mismo patrón de comportamiento climático; así mismo se validó los resultados de las técnicas Machine Learning mediante el índice de siluete, matriz de confusión el cual indica la correcta clasificación de las observaciones en los grupos.
- Los patrones de comportamiento encontrados se los puede explicar mediante la ocurrencia de El Niño Southern Oscillations (ENSO) que tiene un periodo de ocurrencia de aproximadamente 4 años que durante las dos décadas se ha presentado oscilaciones elevadas que ha ocasionado periodos prolongados de fuertes lluvias y sequías.

RECOMENDACIONES

- A posteriori integrar información de otras variables y estaciones meteorológicas a fin de incrementar las observaciones espaciales y mejorar la predicción del comportamiento de temperatura.
- Se recomienda analizar continuamente la información con las etiquetas de tiempo para poder interpretar de mejor manera, para la reducción de la dimensionalidad se recomienda utilizar varios algoritmos seleccionando el que mejor nos dé una respuesta.
- Utilizar la información proporcionada para tomar decisiones tempranas y evitar consecuencias graves, a pesar que los fenómenos no son controlables se puede tomar acciones que eviten un impacto económico ya que se predice que el comportamiento de la temperatura en el rango del periodo 2031-2035 será alta y esto produciría efectos negativos para los seres humanos, flora, fauna, y los glaciares que perderán su capa de hielo.
- Interpretar la información investigada junto a un experto en el área de climatología, existen numerosos factores que los pueden interpretar y emitir soluciones ante el tiempo climático en variación.

BIBLIOGRAFÍA

AGUILERA, P. y SCHMITZ, C. *Manual de prácticas de conservación y gestión en Espacios Naturales Protegidos*. España: Universidad Almería, 2019.

ALPAYDIN, E. *Introduction to Machine Learning*. Tercera ed. London: Massachusetts Institute of Technology. 2014.

ARMANDO, M et al. *Complex land cover change, water and sediment yield in a degraded Andean environment.. Journal of Hydrology*, 2012, p. 6.

ARMENTA, G. *Royecciones climáticas de precipitación y temperatura*. Quito: 2016, p.2.

BENDIX, J. y LAUER, W. *Die niederschlagsjahreszeiten in ecuador und ihre*. Band, Volumen 16, 1992, p.12.

BONET, M. *La Comunidad Geoespacial y el acceso a los datos Oceanográficos*. [Arte] (Instituto Mediterráneo de Estudios Avanzados). 2019, p.10.

BUYTAERT, W. y DE BIEVRE, B. *Water for cities: The impact of climate change and demographic growth in the tropical Andes*. *Water Resources Research*, Volumen 48, 2012, pp. 1-12.

CABALLERO, E. *Microclimas urbanos: la importancia de los materiales*. Primera ed. Santander: Universidad de Cantabria. 2004, p.31.

CACERES, L et al. *Evidencias del cambio climático del Ecuador*. *Bull*, Volumen 3, 1998, p. 549.

CARVAJAL, A. *Los Fundamentos de Lenguaje Phyton*. Chile: 2010, p.21.

CASAS, C. y MARTHA, A. *Meteorología y clima*. Primera ed. Barcelona: UPC. 1999.

CRIPPA, M et al. *Fossil CO2 and GHG emissions of all world countries*. primera ed. Italia: European Union. 2019, p.12.

CROWLEY, T. y NORTH, G. *Abrupt climate change and extinction events in earth history*. *Science*, 240(1), 2018, pp. 996-1002.

CUADRAS, C. *Nuevos Métodos de análisis Multivariante*. Primera ed. Barcelona: CMC. 2018, p.2.

FARFÁN, F. *Agroclimatología del Ecuador*. Quito: Universitaria Abya-Yala. 2018, p.33.

FEDERAL, G. *Cambio Climático ciencia, evidencia y acciones*. [Arte] (Secretaría de Medio Ambiente y Recursos Naturales). 2009, p.14.

GALINDO, E. *Estadística Métodos y Aplicaciones*. [Arte] (Prociencia Editores). 2011, p.5.

GARCÍA, C. *Estadística Descriptiva y Probabilidades para Ingenieros*. Primera ed. Perú: Macro. 2011, p.10.

GARCIA, G. *Investigación comercial*. Segunda ed. Madrid: ESIC. 2005, p.11.

GARCÍA, J. y MOLINA, J. *Ciencia de datos Técnicas analíticas y aprendizaje estadístico*. Primera ed. Bogotá: Altaria S.L. 218, p.26.

GAURISHANKAR, S. *Lecture Notes on Climatology*. Meteorological Training Institute. 2012, p.14.

GONZÁLES, C. y FELPETO, A. *Tratamiento de Datos*. España: Diaz de Santos. 2006.

HANSEN, J. y RUEDY, R. *Global surface temperature change (Geophys)*. 2010.

HERNÁNDEZ, R. y FERNANDEZ, C. *Metodología de la investigación*. Sexta ed. México: mcgraw-hill, 2016, p.12.

HOYER, S. y HAMMAN, J. *Xarray: N-D labeled Arrays and Datasets in Python*. [Arte] (software de investigación abierta). 2007, p.2.

INZUA, J. *Meteorología Descriptiva*. Chile: s.n. 2002.

LOZANO, W. *Clima, hidrología y meteorología.: Para ciencias ambientales e ingeniería*. Primera ed. Bogota: Universidad Piloto de Colombia. 2018, p.12.

MATOS, F. y FORTUNATO, C. *Estadística descriptiva y probabilidad para las ciencias de la información con el uso del spss*. Primera ed. Perú: Asociación de bibliotecólogos del Perú . 2020.

MIRJALILI, V. y RASCHKA, S. *Python Machine Learning*. Segunda ed. España: Marcombo. 2017, p.32.

MONTEALEGRE, J. *Modelo institucional del IDEAM sobre el efecto climático de losfenómenos El Niño y La Niña en Colombia*. Bogota: 2007, p.13.

MÜLLER, A. y GUIDO, S. *Introduction to Machine Learning with Python*. Quinta ed. Estados Unidos: O'Reilly Media. 2016, p.23.

NADAL, I. y MUÑUZURI, V. *Fundamentos de Metereología*. España: Univ Santiago de Compostela. 2006, p.19.

NIKOLAY, A et al. El cambio climático y la conservación de la biodiversidad en el Ecuador. *Cedamaz*, Volumen 1, p. 20.

PATTEN, M. L. & NEWHART, M., 2017. *Understanding Research Methods*. 10 ed. New York: Routledge.

QUEREDA, J. *Curso de Climatología General*. España: Universitat Jaume I. 2005, p.52.

RAY, D. *Estadística para administración y economía. Décima ed. México: cengage learning*. 2008.

RIEBECK, H. *Nasa Earth Observatory*. [en línea] 2005, p.14. Disponible en: <http://earthobservatory.nasa.gov/Features>

RODRIGUEZ, M. *Comparación de métricas de distancia en el algoritmo K-Vecinos Más Cercanos para el problema de Reconocimiento Automático de Dígitos Manuscritos*. [Arte] (Pontificia Universidad Católica de Valparaíso). 2015, p.2.

ROUSSEEUW, P. *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*. *Journal of Computational and Applied Mathematics*, 20(4), 1987, pp. 53-65.

RUSTOM, A. *Estadística descriptiva, probabilidad e inferencia.* Santiago de Chile: Departamento de Economía Agraria .2012.

SAMPIERI, R. *Metodología de la Investigación.* Sexta ed. México: McGRAW-HILL. 2014.

SARKAR, D et al. *Practical Machine Learning with Python.* Primera ed. India: Apress. 2017.

SAROCHAR, H. *Introducción a la Metereología General,* Argentina: Universidad Nacional de la Plata. 2019, p.10.

SERRANO, S. *Análisis estadístico de datos meteorológicos mensuales y diarios para la determinación de variabilidad climática y cambio climático en el distrito metropolitano de quito.* *La Granja*, 2(16), 2012, p. 26.

TORRES, J. *Python Deep Learning: Introducción práctica con Keras y TensorFlow 2.* Primera ed. España: Marcombo. 200, p.24.

VIGUERA, B. y RUTH, M. *El clima, el cambio climático, la vulnerabilidad y acciones contra el cambio climático : Conceptos Básicos.* Costa Rica: Conservación Internacional CATIE. 2017, p18.

VUILLE, M. *Climate Variability in the Andes of Ecuador and Its Relation To Tropical Pacific and Atlantic Sea Surface Temperature Anomalies.* *Journal of climate*, Volumen 13, 2000, pp. 2520-2535.

WU, J. *Advances in K-means Clustering: A Data Mining Thinking.* China: Springer Science & Business Media. 2012, p.14.

XARRAY. *Xarray: matrices y conjuntos de datos etiquetados ND en Python.* [en línea] 2014, Disponible en: <https://xarray.pydata.org/en/latest/index.html>

YAN, Y. & YAN, J., 2018. *Data Science with Anaconda.* Primera ed. Mumbai: Packt.

ZUÑIGA, I. y CRESPO, E., 2015. *Metereología y Climatología.* Primera ed. Madrid: Universidad Nacional de Educación a Distancia.



ANEXO

ANEXO A: OFICIO DE USO DE LA INFORMACIÓN DEL INHAMI



Riobamba, 02 de septiembre del 2021

Señorita

Erika Alexandra Toapanta Yugcha

De mi consideración

Por la presente notifico a usted, que el Grupo de investigación GEEA entrega a usted el **AVAL** y consiente el uso de la información proveniente del INAMHI y satelitales para la realización del trabajo de titulación denominada **"PREDICCIÓN DE LOS PATRONES DE COMPORTAMIENTO DE TEMPERATURA MEDIANTE HERRAMIENTAS ESTADÍSTICAS USADAS EN LOS MODELOS DE CIRCULACIÓN ATMOSFÉRICA 2020-2050"**

Además, la institución patrocinadora declara conocer y aceptar los términos y condiciones previstas para la ejecución del Proyecto, estando conformes con todas aquellas actividades que se prevé realizar.

Sin otro particular, saludo a usted.

Atentamente

Firmado digitalmente por:
CELSO GUILLERMO RECALDE MORENO
CELSO GUILLERMO RECALDE MORENO
Código: 2021.09.02.00401.18
4530
Dr. Celso Guillermo Recalde
DIRECTOR GEEA-FAC. CIENCIAS

Teléfono: 593(03) 2998-200 Ext.2221

geoaespoch@gmail.com



esPOCH

**Dirección de Bibliotecas y
Recursos del Aprendizaje**

**UNIDAD DE PROCESOS TÉCNICOS Y ANÁLISIS BIBLIOGRÁFICO Y
DOCUMENTAL**

REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 16 / 01 / 2023

INFORMACIÓN DEL AUTOR/A (S)
Nombres – Apellidos: ERIKA ALEXANDRA TOAPANTA YUGCHA
INFORMACIÓN INSTITUCIONAL
Facultad: CIENCIAS
Carrera: ESTADISTICA
Título a optar: INGENIERA EN ESTADISTICA INFORMÁTICA
f. Analista de Biblioteca responsable: Ing. CPA. Jhonatan Rodrigo Parreño Uquillas. MBA.


DIRECCION DE BIBLIOTECAS
Y RECURSOS PARA EL APRENDIZAJE
Y LA INVESTIGACION
 Ing. Jhonatan Parreño Uquillas MBA
ANALISTA DE BIBLIOTECA I
2397-DBRA-UPT-2022