



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE CIENCIAS

CARRERA ESTADÍSTICA

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

Trabajo de Titulación

Tipo: Proyecto de Investigación

Presentado para obtener el grado académico de:

INGENIERO EN ESTADÍSTICA INFORMÁTICA

AUTOR:

MIRANDA SAYAY WILSON GEOVANNY

Riobamba – Ecuador

2022



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

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

Trabajo de Titulación

Tipo: Proyecto de Investigación

Presentado para obtener el grado académico de:

INGENIERO EN ESTADÍSTICA INFORMÁTICA

AUTOR: MIRANDA SAYAY WILSON GEOVANNY

DIRECTORA: ING. CRISTINA ESTEFANIA RAMOS ARAUJO MSc.

Riobamba – Ecuador

2022

© 2022, **Wilson Geovanny Miranda Sayay**

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, WILSON GEOVANNY MIRANDA SAYAY, declaro que el presente Trabajo de Titulación es de mi autoría y los resultados del mismo 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, 29 de noviembre de 2022

A handwritten signature in blue ink, appearing to read 'Wilson Geovanny Miranda Sayay', is shown within a rectangular frame. The signature is stylized and somewhat cursive.

Wilson Geovanny Miranda Sayay

0604977215

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 PRECIPITACIÓN MEDIANTE HERRAMIENTAS ESTADÍSTICAS USADAS EN LOS MODELOS DE CIRCULACIÓN ATMOSFÉRICA 2020-2050**, realizado por el señor: **WILSON GEOVANNY MIRANDA SAYAY**, 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 Pérez Londo MSc. PRESIDENTE DEL TRIBUNAL	 _____	2022-11-29
Ing. Cristina Estefanía Ramos Araujo MSc. DIRECTORA DEL TRABAJO DE TITULACIÓN	 _____	2022-11-29
Ing. Johanna Enith Aguilar Reyes MSc. ASESORA DEL TRABAJO DE TITULACIÓN	 _____	2022-11-29

DEDICATORIA

Este trabajo está dedicado a toda mi familia, agradezco por su apoyo incondicional, por cada consejo y confianza que me brindaron en cada momento, en especial a mis padres Resurrección Cunduri y María Sayay, quienes me inculcaron valores y principios, por no dejarme rendirme. A mis hermanas Luz, Elvia y Verónica Miranda por ser los pilares fundamentales para poder cumplir mi meta, usaron su ejemplo y experiencia para guiarme por el camino correcto.

Geovanny

AGRADECIMIENTO

Me gustaría agradecer a todos los profesores y personal de la escuela, quienes imparten conocimientos de semestre a semestre que me ayudaron a convertirme en una persona y un profesional. Además, quisiera agradecer a la Ing. Cristina Ramos, la tutora de mi tesis y al grupo de investigación GEAA, quienes me dieron confianza para poder realizar el trabajo de titulación, también a la Ing. Johanna Aguilar, Miembro de mi trabajo de grado en la institución y docente de diversas disciplinas, quienes me han guiado de la mejor manera por muchos años, el camino como estudiante y en este trabajo.

Geovanny

TABLA DE CONTENIDO

ÍNDICE DE TABLAS.....	x
ÍNDICE DE FIGURAS.....	xi
ÍNDICE DE GRÁFICAS.....	xii
ÍNDICE DE ECUACIONES	xiii
ÍNDICE DE ANEXOS	xiv
RESUMEN.....	xiii
SUMMARY	xiv
INTRODUCCIÓN	1

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL.....	7
1.1. Bases conceptuales	7
1.1.1. <i>Análisis exploratorio de datos</i>	7
1.1.1.1. <i>Estadística</i>	7
1.1.2. <i>Medidas de Tendencia Central</i>	7
1.1.2.1. <i>Media Aritmética</i>.....	7
1.1.2.2. <i>Mediana</i>	7
1.1.2.3. <i>Medidas de Dispersión</i>.....	8
1.1.2.4. <i>Desviación Estándar</i>	8
1.1.2.5. <i>Varianza</i>	8
1.1.3. <i>Técnicas Machine Learning</i>	8
1.1.4. <i>Técnicas estadísticas Multivariantes</i>	9
1.1.5. <i>Reducción de dimensionalidad</i>	9
1.1.5.1. <i>Definición de Reducción de dimensionalidad</i>.....	9
1.1.5.2. <i>Componentes Principales</i>.....	10
1.1.5.3. <i>Variabilidad explicada por los componentes</i>.....	11
1.1.5.4. <i>Técnica de Incrustación de vecinos estocásticos distribuidos en t (t-SNE)</i>.....	11
1.1.5.5. <i>Funcionamiento de t-SNE</i>	11
1.1.5.6. <i>Pasos básicos de cálculo de t-SNE</i>	12
1.1.5.7. UMAP: Proyección y Aproximación de Múltiple Uniforme para Reducción de Dimensiones.....	13
1.1.5.8. <i>Base matemática de UMAP</i>	14

1.1.6.	<i>Técnicas Clustering</i>	15
1.1.7.	<i>Los principales algoritmos de Clustering</i>	15
1.1.7.1.	<i>Diagrama de dispersión algoritmo UMAP agrupación K-Means</i>	15
1.1.7.2.	<i>Diagrama de dispersión algoritmo UMAP agrupación DBSCAN</i>	17
1.1.7.3.	<i>El algoritmo DBSCAN se puede abstraer en los siguientes pasos</i>	17
1.1.7.4.	<i>Modelo jerárquico distancia euclídea enlace Ward-Algoritmo UMAP</i>	18
1.1.7.5.	<i>Optimizando UMAP- métrica chebyshev</i>	19
1.1.8.	<i>Instrumentos de procesamiento</i>	20
1.1.8.1.	<i>Anaconda</i>	20
1.1.8.2.	<i>Jupyter Notebook</i>	20
1.1.8.3.	<i>Python</i>	20
1.1.9.	<i>Modelo de Circulación Atmosférico</i>	20
1.1.9.1.	<i>Anomalías Climáticas</i>	21
1.1.9.2.	<i>Anomalía</i>	22
1.1.9.3.	<i>¿Por qué son útiles?</i>	22
1.2.	Bases teóricas	23
1.2.1.	<i>Sistema Climático</i>	23
1.2.2.	<i>Cambio Climático</i>	23
1.2.3.	<i>Dinámica Atmosférica</i>	24
1.2.3.1.	<i>La humedad</i>	24
1.2.3.2.	<i>Humedad relativa</i>	25
1.2.4.	<i>La circulación general atmosférica</i>	25
1.2.5.	<i>Precipitación</i>	26
1.2.5.1.	<i>Causas</i>	26
1.2.5.2.	<i>Mecanismo</i>	27
1.2.5.3.	<i>Latitud</i>	27
1.2.5.4.	<i>Longitud</i>	27
1.2.5.5.	<i>Tiempo</i>	27
1.2.6.	<i>Condensación</i>	28
1.2.6.1.	<i>Coalescencia</i>	28
1.2.7.	<i>Precipitación en el Ecuador</i>	29
1.2.7.1.	<i>Humedad (relativa) en el ecuador</i>	30
1.2.8.	<i>Patrones climáticos que afectan el clima en el Ecuador (ENSO)</i>	31
1.2.8.1.	<i>Anticiclón del pacífico</i>	32

CAPÍTULO II

2.	MARCO METODOLÓGICO	33
2.1.	Tipo de la Investigación	33
2.2.	Planteamiento del problema	33
2.3.	Revisión de la literatura y desarrollo del marco teórico	33
2.4.	Visualización del alcance del estudio	34
2.5.	Desarrollo del diseño de investigación	34
2.6.	Análisis de datos.....	34
2.7.	Evaluación de los reportes	34
2.8.	Elaboración del reporte de resultados	35
2.9.	Diseño de la investigación	35
2.10.	Localización de estudio	35
2.11.	Población de estudio	35
2.12.	Método de muestreo	36
2.13.	Tamaño de la muestra	36
2.14.	Técnica de recolección de datos.....	36
2.15.	Identificación de variables	36
2.16.	Modelo estadístico.....	36
2.17.	Variables en estudio.....	37
2.18.	Operacionalización de variables	37

CAPÍTULO III

3.	MARCO DE RESULTADOS Y DISCUSIÓN DE RESULTADOS	38
3.1.	Análisis Descriptivo	38
3.1.1.	<i>Variables cuantitativas</i>	38
3.1.2.	<i>Descripción de datos</i>	38
3.1.3.	<i>Estadística Descriptiva</i>	39
3.1.4.	<i>Exploración de las anomalías</i>	41
3.1.5.	<i>Matriz bidimensional</i>	42
3.1.6.	<i>Reducción de dimensionalidad</i>	42
3.1.7.	<i>Componentes principales</i>	43
3.1.8.	<i>Técnica de Incrustación de vecinos estocásticos distribuidos en t (t-SNE)</i>	45
3.1.9.	<i>UMAP: Proyección y Aproximación de Múltiple Uniforme</i>	45
3.1.10.	<i>Aplicación de Algoritmos de Clustering</i>	46
3.1.11.	<i>Selección del número óptimo de Clusters</i>	46
3.2.	Algoritmo UMAP (métrica=euclídean) agrupación DBSCAN	47

3.2.1.	<i>Matriz de confusión</i>	48
3.2.2.	<i>Representación de resultados por días del algoritmo DBSCAN</i>	48
3.2.3.	<i>Mapa de calor de patrones de comportamiento</i>	50
3.2.4.	<i>Algoritmo UMAP (métrica -euclídean) agrupación K-Means</i>	53
3.2.5.	<i>Algoritmo UMAP (métrica euclídean) agrupación Agglomerative Clustering</i>	53
3.2.6.	<i>Algoritmo UMAP (métrica-chebyshev) agrupación -K-Means</i>	54
3.2.7.	<i>Algoritmo UMAP (métrica-chebyshev) agrupación Agglomerative Clustering</i>	55
3.2.8.	<i>Algoritmo UMAP (métrica-euclídea) agrupación DBSCAN</i>	56
3.2.9.	<i>Validación de Modelo</i>	56
CONCLUSIONES		59
RECOMENDACIONES		60
BIBLOGRAFÍA		
ANEXO		

ÍNDICE DE TABLAS

Tabla 1-1:	Ventajas y Desventajas de las Técnicas Clusters	19
Tabla 1-2:	Operacionalización de variables.....	37
Tabla 1-3:	Tabla de contingencia de la variable Precipitación Tridimensional	38
Tabla 2-3:	Estadística Descriptiva	39
Tabla 3-3:	Dimensiones de la matriz Bidimensional de Precipitación	42
Tabla 4-3:	Rango de años por cada grupo.....	48
Tabla 5-3:	Matriz de confusión.....	48
Tabla 6-3:	Rango de años para los grupos de clúster.....	48
Tabla 7-3:	Índice de silhouette.....	56
Tabla 8-3:	Precipitaciones registradas a nivel nacional mes de enero y febrero de 2021	57
Tabla 9-3:	Precipitaciones registradas a nivel nacional marzo - abril de 2021	57
Tabla 10-3:	Precipitaciones registradas a nivel nacional mayo- agosto de 2021	58
Tabla 11-3:	Precipitaciones registradas a nivel nacional septiembre- diciembre de 2021.....	58

ÍNDICE DE FIGURAS

Figura 1-1:	Técnicas Machine Learning	9
Figura 2-1:	Reducción de dimensionalidad	10
Figura 3-1:	Gráfico de t-SNE.....	12
Figura 4-1:	Medición de similitudes por pares en el espacio de alta dimensión.....	12
Figura 5-1:	Distribución t normal vs. Student	13
Figura 6-1:	Proyección UMAP 3d.....	14
Figura 7-1:	Técnicas Clustering.....	15
Figura 8-1:	Agrupación K-means.	16
Figura 9-1:	Agrupación DBSCAN	17
Figura 10-1:	Análisis de conglomerados con el método de Ward y distancia Euclídea.....	18
Figura 11-1:	Datos Tridimensional.....	21
Figura 12-1:	Anomalías de Climatología de 19950 a 2000	21
Figura 13-1:	Anomalías positivas y negativas	22
Figura 14-1:	Cuadro de valores del índice normalizado de precipitación.	23
Figura 15-1:	Dinámica Atmosférica	24
Figura 16-1:	Circulación General de la Atmosfera.....	26
Figura 17-1:	Precipitación	26
Figura 18-1:	Coalescencia	28
Figura 19-1:	Precipitación Anual.....	29
Figura 20-1:	Altitud vs Humedad relativa Anual	30
Figura 21-1:	Humedad relativa anual-Ecuador.....	31
Figura 22-1:	Corriente del Niño.....	32
Figura 1-2:	Proceso cuantitativo	33
Figura 2-2:	Mapa del Ecuador	35

ÍNDICE DE GRÁFICAS

Gráfico 1-3:	Histograma de frecuencias de promedio de anomalías por filas	40
Gráfico 2-3:	Gráfico de líneas promedio de anomalías por año	41
Gráfico 3-3:	Promedio de anomalías por mes.....	42
Gráfico 4-3:	Componentes Principales.	43
Gráfico 5-3:	Porcentaje de varianza Explicada.....	44
Gráfico 6-3:	Diagrama de dispersión de los 2 primeros componentes principales por año. ...	44
Gráfico 7-3:	Técnica de Incrustación de vecinos estocásticos t-SNE por año.....	45
Gráfico 8-3:	Diagrama de dispersión algoritmo UMAP por años.	46
Gráfico 9-3:	Gráfica de numero de clústeres.	46
Gráfico 10-3:	Diagrama de dispersión UMAP agrupación DBSCAN.....	47
Gráfico 11-3:	Frecuencia de días en cada uno de los grupos	49
Gráfico 12-3:	Mapa de calor de Patrones de comportamiento.....	50
Gráfico 13-3:	Mapa de calor Patrones de comportamiento.	50
Gráfico 14-3:	Mapa de calor Patrones de comportamiento.	51
Gráfico 15-3:	Mapa de calor Patrones de comportamiento.	51
Gráfico 16-3:	Mapa de calor Patrones de comportamiento.	52
Gráfico 17-3:	Diagrama de dispersión algoritmo UMAP agrupación K-Means.	53
Gráfico 18-3:	Diagrama de dispersión algoritmo UMAP agrupación Agglomerative Clustering.	53
Gráfico 19-3:	Diagrama de dispersión UMAP agrupación K-Means por categoría.	54
Gráfico 20-3:	Diagrama de dispersión UMAP agrupación K-Means por años.	54
Gráfico 21-3:	Diagrama de dispersión UMAP agrupación Agglomerative Clustering	55
Gráfico 22-3:	Diagrama de dispersión UMAP agrupación DBSCAN.....	56
Gráfico 23-3:	Índice de silhouette.....	57

INDICE DE ECUACIONES

Ecuación 1-1:	Media Aritmética	7
Ecuación 2-1:	Mediana	8
Ecuación 3-1:	Desviación Estándar	8
Ecuación 4-1:	Varianza.....	8
Ecuación 5-1:	Componentes Principales.....	10
Ecuación 6-1:	Transformación Lineal.....	11
Ecuación 7-1:	Variabilidad Explicada por los Componentes	11
Ecuación 8-1:	Agrupación de K-Means.....	16
Ecuación 9-1:	Distancia Euclídea enlace Ward.....	19
Ecuación 10-1:	Métrica Chebyshey	19
Ecuación 11-1:	Humedad Relativa	25

ÍNDICE DE ANEXOS

ANEXO A: AVAL DE GRUPO DE INVESTIGACIÓN GEAA

RESUMEN

El objetivo de esta investigación fue identificar patrones de comportamiento de precipitación del periodo 2020-2050 mediante herramientas estadísticas, técnicas machine learning y validación de modelos mediante datos observados y predichos. La base de datos de precipitación se obtuvo de los modelos atmosféricos del Centro Nacional de Investigaciones Meteorológicas (CNRM), que se encuentra en el proyecto Inter comparaciones de Modelos Acoplados Fase 6 (CMIP6), los datos de precipitación fueron del periodo 2020-2050 recolectadas a escala diaria, el cual para este análisis de datos se utilizó el lenguaje de programación Python, para la elaboración de histogramas de frecuencia y técnicas multivariadas para la reducción de dimensionalidad; se aplicó diferentes técnicas Machine Learning para identificar patrones de comportamiento como: K-Means, Agrupamiento Espacial Basado en Densidad de Aplicaciones con Ruido (DBSCAN) y Agrupación Aglomerativo, donde la técnica (DBSCAN) identificó 4 clústeres de patrones de comportamiento, el cual para validar dicho modelo se utilizó la métrica de precisión de Silhouette y la información de precipitaciones registradas en el Ecuador del año 2021. Como resultado de este análisis arrojó como la mejor técnica para la reducción de dimensionalidad e identificación de patrones de comportamiento de precipitación fue el algoritmo con métrica euclídea (DBSCAN), con una efectividad del 70.92% mayor que las demás técnicas empleadas, los patrones de comportamiento identificados con la técnica ya mencionada fueron similares a los datos predichos. Se recomienda experimentar con varias técnicas machine learning para observar cuál de estas técnicas es el adecuado para la base de datos en estudio, es importante acudir donde un experto en el área de la climatología para una mejor interpretación y comprensión de los eventos climatológicos.

Palabras clave: <PRECIPITACIÓN>, <REDUCCIÓN DE DIMENSIONALIDAD>, <TÉCNICAS MACHINE LEARNING>, <CMIP6>, <VALIDACIÓN DE MODELOS >, <DBSCAN>.

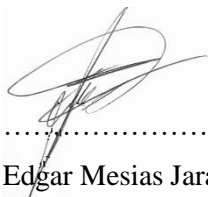


2390-DBRA-UPT-2022

SUMMARY

The objective of this research was to identify behavior patterns of precipitation for the 2020-2050 period through statistical tools, machine learning techniques and model validation using observed and predicted data. The precipitation database was obtained from the atmospheric models of the National Center for Meteorological Research (CNRM), which is in the project Intercomparisons of Coupled Models Phase 6 (CMIP6), the precipitation data was collected from the period 2020-2050 on a daily scale, for this data analysis the Python programming language was used to elaborate frequency histograms and multivariate techniques for dimensionality reduction; different Machine Learning techniques were applied to identify behavior patterns such as: K-Means, Density-based Spatial Clustering of Applications with Noise (DBSCAN) and Agglomerative Grouping, where the technique (DBSCAN) identified 4 clusters of behavior patterns, in order to validate said model, the Silhouette precision metric and the precipitation information recorded in Ecuador for the year 2021 were used. As a result of this analysis, the best technique for the reduction of dimensionality and identification of patterns of precipitation behavior was the algorithm with Euclidean metrics (DBSCAN), with an effectiveness of 70.92% greater than the other techniques used, the behavior patterns identified with the aforementioned technique were similar to the predicted data. It is recommended to experiment with various machine learning techniques to see which of these techniques is appropriate for the database under study. It is important to go to an expert in the area of climatology for a better interpretation and understanding of weather events.

Keywords: <PRECIPITATION>, <DIMENSIONALITY REDUCTION>, <MACHINE LEARNING TECHNIQUES>, <CMIP6>, <MODEL VALIDATION>, <DBSCAN>.



.....
Edgar Mesias Jaramillo Moyano

0603497397

INTRODUCCIÓN

A pesar de la incertidumbre, el calentamiento global y el progreso de la población, la brecha Norte-Sur o la universalización, ha sido un tema de gran preocupación social, pues su principal responsabilidad son las actividades humanas, por el impacto ambiental como también la salud del pueblo, especialmente por ser un país en desarrollo, otro problema son las tensiones entre países subdesarrollados, ricos y pobres, justicia o injusticia, reparto de cargas y esfuerzos desarrolladores de recursos y sociedades explotadas, o simplemente, control del uso de combustibles y de la tierra, estrictamente en otras palabras, ser responsables de esta y de las generaciones futuras, su bienestar debe mejorarse con el desarrollo de tecnología (Fernández , 2013, p. 72).

Las Ciencias Atmosféricas se han vuelto muy importante en los últimos años porque estudia la dinámica atmosférica. El cambio climático se ve reflejado en sequías, inundaciones y fenómenos meteorológicos extremos que han afectado a diversos sectores productivos y económicos del país. También a principios de 2008, Ecuador sufrió una de las peores lluvias e inundaciones de la historia, que afectó a 13 provincias, y los sectores de agricultura e infraestructura sufrieron pérdidas considerables, especialmente en las zonas rurales, así como en los agricultores, terratenientes y fincas. Los trabajadores han sido los más afectados, el cambio climático se ha reflejado en variaciones de la precipitación en la distribución temporal y espacial, así mismo, extensión constante de la temperatura y un acrecentamiento en la periodicidad y severidad de los fenómenos climáticos. Los glaciares retroceden y la altitud de los océanos como también de mares incrementan (Armijos & Loján , 2010, p. 18).

Además, se han logrado avances significativos en la predicción de las precipitaciones, que desempeña un papel muy importante en la vida humana. La predicción precisa a largo plazo de precipitaciones de invierno y verano es un desafío, varios centros meteorológicos están implementado inteligencia artificial para mejorar los pronósticos numéricos. La Administración Nacional Oceánica y Atmosférica (NOAA), la agencia líder del gobierno de los EE. UU., tiene la intención de expandir el uso de la inteligencia artificial en cada misión de la NOAA para mejorar la eficiencia, la eficacia y la coordinación del desarrollo y la oportunidad de la ciencia, los productos y servicios (Ankush, et al., 2020, p. 2).

El Ecuador se divide en cuatro regiones: costa, sierra, oriente y galápagos, en cada territorio podemos sentir un clima único dependiendo de ubicación geográfica, la región costa se caracteriza por tener un clima tropical, seco y húmedo. La región interandina suele tener regiones frías, templadas y cálidas, dependiendo de la altura relativa al nivel del mar. En la Amazonía, el clima

es más cálido y húmedo. Finalmente, en la región de Galápagos, el clima se define como una zona templada y seca (Pourrut, 1987, p. 13).

El trabajo reciente que muestra la viabilidad de usar el aprendizaje automático para la predicción del clima, los modelos de circulación atmosférica y la combinación de técnicas de inteligencia artificial (IA) y la ciencia meteorológica ha llevado a los investigadores a implementar métodos de IA para predecir y clasificar patrones climáticos. Las técnicas estadísticas ayudan a comprender el comportamiento de ciertos fenómenos y tomar decisiones correctas para resolver problemas. La intención de la exploración, es utilizar herramientas estadísticas innovadoras, para observar el comportamiento de patrones de precipitación. Las diversas variables, que trasladan a cambios en la precipitación que se muestra en el Ecuador, tratando de inferir más escenarios precisos, así poder contrastar la diferencia que existe en patrones de comportamiento de precipitación presentes y futuro. Por lo descrito, esta investigación tiene como objetivo identificar los patrones de comportamiento de precipitación del periodo 2020-2050.

Antecedentes

El impacto observado en el Ecuador se ha reflejado en la percepción biofísica del territorio, como el deterioro de los glaciales de los Andes, aumento de temperatura, inundaciones, sequías, acrecentamiento del nivel del mar, impactos negativos en las actividades productoras y la biodiversidad, etc. Asimismo, la ocurrencia de fenómenos climáticos como El Niño Southern Oscillation (ENSO). Hay dos escenarios en Ecuador: ENSO (fase cálida) y ENOS (fase fría), donde las precipitaciones se incrementan y se disminuyen respectivamente. Las consecuencias observadas se reflejan en la extensión de la temperatura como también de las precipitaciones (Aguilar, 2020, pp. 74-75).

Según la división gubernamental sobre el calentamiento global, el cambio climático conducirá a una extensión progresiva de la temperatura media del área terrestre y los océanos, cambios en los patrones de comportamiento de la precipitación, cambios en la intensidad, frecuencia de los eventos climáticos, eventos extremos y escalada del nivel del mar. Se estima que, a finales del siglo XXI, la temperatura de la superficie terrestre puede subir entre 2,6 y 4,8 grados centígrados, y el acrecentamiento medio del nivel del mar oscila entre 45, y 82 centímetros. Las precipitaciones en las extensiones altas pueden aumentar, mientras que en las regiones subtropicales pueden disminuir (Botero, 2015, p. 13).

De acuerdo con el informe del Panel Internacional sobre el Cambio Climático América Latina es una de las regiones vulnerables a esos cambios, y sufrirá pérdidas económicas, humanas,

infraestructuras, agrícola y ambientales. Según el Panel Intergubernamental de Cambio Climático (PICC) afirma que la concentración de dióxido de carbono en la atmósfera ha aumentado a 379 partes por millón (un número nunca alcanzado en los últimos 650 mil años de historia de la tierra), como también la concentración de metano y óxido nitroso, el océano es 17 cm más alto que a principios del siglo XX y la temperatura 0. 7° más cálida que en 1850. A consecuencia de ello han aumentado las precipitaciones en ciertas regiones, hay más intensas y prolongadas sequías en amplias zonas de los trópicos y subtropicos, la frecuencia de las precipitaciones fuertes ha aumentado en varios países (OG, 2007, p. 22).

ENSO se encuentra establecido en el Pacífico afectando severamente a los países de América Latina. Las consecuencias más evidentes son las sequías, como en el caso de Colombia, Venezuela, México y Centroamérica; fuertes lluvias en Argentina, el norte de Perú y sur de Chile y el incremento de incendios forestales en zonas como Brasil. Las precipitaciones, inferiores a lo normal, ya han provocado sequía en zonas de Venezuela y norte de Colombia e inundaciones en Perú, principalmente en el sector agrícola que generan aumentos significativos en los precios de los alimentos (Virginia, 1998, p. 24).

Planteamiento del problema

- **Enunciado del problema**

Las inundaciones y sequías han provocado pérdidas materiales, cultivos, pérdidas humanas que afectan directamente la economía del país. De acuerdo al Servicio Nacional de Gestión de Riesgos en los últimos seis meses, las consecuencias de las lluvias en el Ecuador dejan 57 muertos, 110 heridos, 31.000 afectados, 5.455 hectáreas agrícolas afectados y destruidas, así como también 7. 176 viviendas, centros de salud y de educación dañados parcial o totalmente. Las lluvias afectaron a 23 de las 24 provincias de Ecuador, siendo la excepción Galápagos, un archipiélago a 1.000 kilómetros de la costa (Ocles, 2022, p. 4).

Se espera que el cambio climático tenga impactos significativos en la biodiversidad, con impactos directos en las poblaciones y comunidades dependientes de la agricultura, la pesca, el turismo, flora y fauna entre otras actividades económicas que de alguna manera requieren la conservación de los recursos biológicos y ecosistémicos (Botero, 2015, p. 11).

Ecuador presento pérdidas económicas de 424 millones de dólares en la agricultura tomando como referencia a 505.000 hectáreas de terrenos, el 66% de las hectáreas pertenecen a la región

Costa etas información se presentó en los últimos 17 años, mientras que en el 2020 se han perdido 27.000 hectáreas por los incendios forestales (Loaiza, 2021, p. 5).

El aumento del nivel del mar resultará en una pérdida del 1 % al 2% de los ecosistemas de manglares por año. Esto a su vez afectará la dinámica poblacional de ciertos peces, moluscos, anfibios y mamíferos acuáticos como las ballenas. De 1993 a 2018, el nivel del mar en Ecuador aumentó 2,6 mm por año, según la agencia de la ONU IPCC (Pesantes, 2020, p. 3).

Además, los seres humanos sono susceptibles a las enfermedades contraídas por efectos en el cambio climático como variaciones extremas en precipitación y temperatura. Estos incluyen enfermedades comunes transmitidas por vectores como la malaria y el dengue, pero también otras causas importantes de muerte como la desnutrición y la diarrea. El cambio climático sigue aumentado la carga mundial de morbilidad (OPS, 2019, p. 2).

- **Formulación (Incógnita)**

- ¿Existen cambios significativos de precipitación en el Ecuador durante el periodo 2020 a 2050?
- ¿Qué factores define el comportamiento climático en el Ecuador?
- ¿Cuál es la mejor técnica Machine learning que ha producido mejores resultados para determinar patrones de comportamiento?
- ¿Qué precisión tiene la técnica Machine Learning para alcanzar el objetivo de la investigación?

Justificación

Esta investigación se centra en analizar la alteración climática que se dará en años posteriores con aumento significativo de la precipitación, se prevé la existencia de fechas con aumento de lluvias que ocasionan inundaciones en diferentes sectores del Ecuador debido al fenómeno ENSO, cerca de 7 millones de personas se han visto afectadas. El 60% de la población ecuatoriana, ha experimentado cambios en sus condiciones de vida, en 1997-1998 el fenómeno ENSO ha tenido un gran impacto principalmente en la región costa que ocupa alrededor del 40% del país y alberga a 6,7 millones de personas. Un total de 292 personas fallecieron, 162 resultaron heridas y 40 desaparecidos por causas relacionadas al fenómeno (OPS, 2019, p. 4).

En el Ecuador se estima en US\$ 2.869,3 millones de pérdidas económica; de ellos, US\$ 783,2 millones (27%) corresponden a daños directos y US\$ 2.086,1 millones (73%) a daños indirectos, según el reporte de la Defensa Civil (CEPAL, 2005, p. 197). Las temporadas de sequías ocasionan pérdidas económicas. El cálculo de los daños como consecuencia alcanzó un total de US\$ 640,6 millones. De esa cantidad, US\$ 533,9 millones (83%) fueron por daños directos y los restantes US\$ 106,7 millones (17%) por daños indirectos (CEPAL, 2005, p. 187).

Como expresa Antonio Almeida director del pronóstico y alertas hidrometeorológicas del Inamhi expone que la inteligencia artificial (IA) tiene mucho potencial para el monitoreo del tiempo y las predicciones a largo plazo, afirma que las técnicas Machine learning o Deep Learning puedan utilizarse para cualquier aplicación (tendencias de variables meteorológicas y series de tiempo). Estas técnicas manejan grandes volúmenes de datos, que son de utilidad para estudiar el comportamiento de la dinámica atmosférica y facilitar la elaboración de pronósticos (Lara, 2022, p. 34).

El propósito de la investigación es identificar patrones de comportamiento a futuro, mediante técnicas Machine Learning que permitan establecer temporadas climáticas, identificar variaciones en la precipitación y alertar a las personas dedicadas a la agricultura y pesca para evitar pérdidas sustanciales, costos elevados en los productos que se comercializan en los mercados ecuatorianos.

OBJETIVOS

Objetivo General

- Realizar predicción de patrones de comportamiento de precipitación mediante herramientas estadísticas usadas en los modelos de circulación atmosférica 2020-2050.

Objetivos Específicos

- Explorar los datos obtenidos de los modelos atmosféricos y estaciones meteorológicas de INAMHI.
- Predecir patrones de comportamiento de precipitación mediante técnicas Machine Learning.
- Validación del 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. *Análisis exploratorio de datos*

1.1.1.1. *Estadística*

El campo de la estadística trata de la recolección, presentación, procesamiento, análisis y uso de datos para tomar decisiones, solucionar problemas, diseñar productos y procesos (Matus, 2010, p. 1).

1.1.2. *Medidas de Tendencia Central*

Si un conjunto de datos se ordena de forma ascendente o descendente, cualquier medida que esté situada en el centro de dicho conjunto se le llama medida de centralización. También se le puede llamar promedio (Martínez, 2020, p. 129).

1.1.2.1. *Media Aritmética*

La meda aritmética o simplemente media, que se denota por \bar{X} , es el número obtenido al dividir sumar todos los productos de la variable divididos por el número general de observaciones, y se define por las siguientes expresiones (Muñoz, 2000, p. 8).

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \qquad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \qquad (1-1)$$

1.1.2.2. *Mediana*

La mediana de la sucesión de datos ordenados en orden ascendente o descendente es el dato central; Si el número de datos es impar; y es el promedio de los dos datos centrales, si el número de datos es impar (Martínez, 2020, p. 141).

$$\tilde{x} = \begin{cases} x_{(n+1)/2} & \text{si } n \text{ es impar} \\ \frac{1}{2} \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right) & \text{si } n \text{ es par} \end{cases} \quad (2-1)$$

1.1.2.3. Medidas de Dispersión

Intentan medir qué tan dispersa es una variable estadística con respecto a la medida de posición o tendencia central, dada la representatividad de la escala de posición. Cuanto mayor sea la dispersión, menor será la representación de la posición y viceversa (Muñoz, 2000, p. 17).

1.1.2.4. Desviación Estándar

Es la media de los cuadrados de las desviaciones, como es la desviación estándar es la raíz cuadrada de la varianza (Smith & Pierce, 1991, p. 52).

$$S = \sqrt{\frac{\sum_{i=1}^n d_i}{n-1}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (3-1)$$

1.1.2.5. Varianza

Mide menor o mayor dispersión de los valores de una variable con respecto a la media aritmética. Cuanto mayor es la varianza, mayor es la dispersión y, por tanto, menos representativa la media (Muñoz, 2000, p. 18).

$$S^2 = \frac{\sum_{i=1}^n d_i}{n-1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (4-1)$$

1.1.3. Técnicas Machine Learning

El aprendizaje automático, o más conocido como Machine Learning, es la ciencia de hacer que las computadoras "aprendan" de los datos. El campo del aprendizaje automático ya no se trata de programar cada solución específica para cada requisito paso a paso, como lo hacen los métodos de programación tradicionales, sino que se centra en desarrollar algoritmos generales que pueden extraer patrones de diferentes tipos de datos. Existen algunos tipos de Machine Learning como son: aprendizaje supervisado, aprendizaje no supervisado, aprendizaje semi-supervisado y aprendizaje por esfuerzo (Bobadilla, 2021, p. 14).

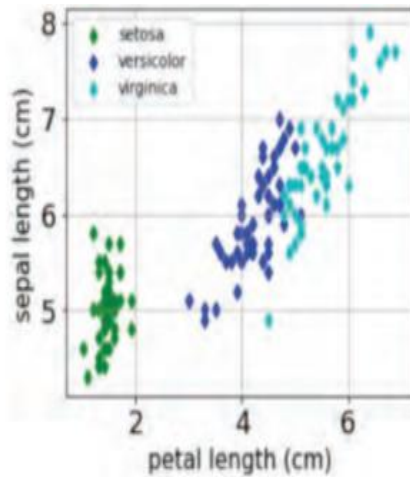


Figura 1-1: Técnicas Machine Learning

Fuente: Bobadilla Jesús, 2021.

1.1.4. Técnicas estadísticas Multivariantes

El análisis multivariada es una parte del análisis, representación e interpretación de datos estadísticos y de investigación de datos obtenidos al observar más de una variable estadística en una muestra individual. Las variables observadas son homogéneas e interrelacionadas, y una no domina a la otra. La información estadística en el análisis multivariante es de naturaleza multidimensional, por lo que la geometría, la aritmética matricial y la distribución multivariante juegan un papel esencial (Cuadras, 2018, p. 13).

La investigación multivariado es un arreglo bidimensional de número de datos, pero en Análisis Multivariante, la información de entrada generalmente consiste en una matriz de distancia o similitud, que mide el grado de diferencia entre los sujetos. Se emprenderá con una técnica basada en matrices de datos $n \times p$, en que n es el dígito de individuos y p es el número de variables (Cuadras, 2018, p. 13).

1.1.5. Reducción de dimensionalidad

1.1.5.1. Definición de Reducción de dimensionalidad

A menudo se usa como un paso de preprocesamiento en otro tipo de trabajo de aprendizaje automático, principalmente en clasificación o regresión. Muchas situaciones del mundo real proporcionan datos escasos o datos que a menudo proporcionan muy poca información. Lo que estamos haciendo aquí es transformar datos multidimensionales en una dimensión para cada característica: para cada elemento, que está muy disperso, en información multidimensional cada

vez más densa. La siguiente figura muestra una imagen fuerte y diferentes versiones de la imagen obtenida aplicando diferentes reducciones de dimensionalidad (Bobadilla, 2021, p. 18).



Figura 2-1: Reducción de dimensionalidad

Fuente: Bobadilla Jesús, 2021.

El aprendizaje semi-supervisado se ocupa de conjuntos de datos en los que una parte de los datos está etiquetada y el resto no. Por lo general, el número de muestras etiquetadas es mucho menor que el de las no etiquetadas, la totalidad de algoritmos de aprendizaje semi-supervisado son una mezcla de algoritmos de aprendizaje supervisados y no supervisados (Bobadilla, 2021, p. 18).

1.1.5.2. Componentes Principales

Un problema central en el estudio de antecedentes multivariados es la disminución de la dimensionalidad, si los valores de p variables pueden describirse con exactitud mediante un pequeño subconjunto de ellas, $r < p$, entonces la dimensionalidad del problema se reduce, pero se pierde una pequeña parte de la información. El objetivo del análisis de componentes principales es, dadas n observaciones de p variables, analizar si es posible representar apropiadamente esta información con un número menor de variables construidas como combinaciones lineales de las variables originales (Peña, 2002, p. 137).

1. Permite que las observaciones de espacios p -dimensionales generales se representen de manera óptima en espacios de pequeña dimensión. En este sentido, los componentes principales son el primer paso para identificar variables latentes o no observadas que pueden generar variabilidad en los datos (Peña, 2002, p. 137).
2. Permite la transformación de variables originales normalmente correlacionadas en nuevas variables no correlacionadas y facilita la interpretación de los datos (Peña, 2002, p. 137).

Las componentes principales son las variables compuestas (Cuadras, 2018, p. 77).

$$Y_1 = Xt_1, Y_2 = Xt_2, \dots, Y_p = Xt_p \quad (5-1)$$

Si $T=[t_1, t_2, t_3, \dots, t_p]$ es una matriz $p \times p$ cuyas columnas son vectores que definen los componentes principales, seguidos de la transformación lineal (Cuadras, 2018, p. 78).

$$X \rightarrow Y \quad ; \quad X \rightarrow Y \quad (6-1)$$

1.1.5.3. Variabilidad explicada por los componentes

En aplicaciones, se espera que el primer componente explique un alto porcentaje de la variación total. Un breve ejemplo, si $m = 2 < p$; y $P_2 = 90 \%$; los dos primeros componentes explican gran parte de la variación de las variables. Entonces lograremos sustituir X_1, X_2, \dots, X_p por las componentes principales Y_1, Y_2 . En muchas aplicaciones, estos componentes tienen interpretaciones experimentales (Cuadras, 2018, p. 79).

$$P_m = 100 \frac{\lambda_1 + \dots + \lambda_m}{\lambda_1 + \dots + \lambda_p} \quad (7-1)$$

1.1.5.4. Técnica de Incrustación de vecinos estocásticos distribuidos en t (t -SNE)

Es una técnica no lineal no supervisada que se utiliza principalmente para la exploración y visualización de datos de alta dimensión. En pocas palabras, t -SNE le brinda una sensación o intuición sobre cómo se organizan sus datos en un espacio de alta dimensión, este método se utiliza en investigaciones de cáncer, investigación climática, seguridad informática, entre otros. Fue desarrollado en 2008 por Laurens van der Maaten y Geoffrey Hinton (Violante, 2018, p.6).

1.1.5.5. Funcionamiento de t -SNE

El algoritmo t -SNE calcula una medida de similitud entre pares de instancias en un espacio de alta y baja dimensión. Luego, intenta optimizar estas dos medidas de similitud usando una función de costo (Violante, 2018, p.6).

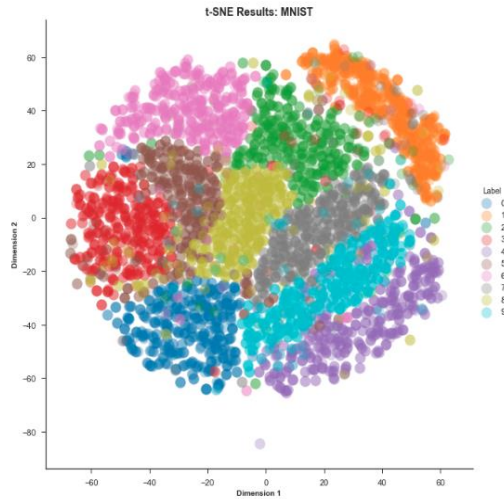


Figura 3-1: Gráfico de t-SNE

Fuente: Violante Andre, 2018.

1.1.5.6. Pasos básicos de cálculo de t-SNE

Paso 1: Mide la similitud entre puntos en un espacio de alta dimensión. Piense en un montón de puntos de datos dispersos en un espacio bidimensional (Figura 4-1). Para cada punto de datos (x_i), centraremos la distribución gaussiana en ese punto. Luego se mide la densidad de todos los puntos (x_j) bajo esta distribución gaussiana. Luego se normaliza todos los puntos nuevamente. Esto nos da un conjunto de probabilidades (P_{ij}) para todos los puntos. Estas probabilidades son proporcionales a la similitud, significa que, si los puntos de datos x_1 y x_2 tienen valores iguales bajo este círculo gaussiano, entonces sus proporciones y similitudes son iguales y, por lo tanto, tienen similitud local en la estructura de este espacio de alta dimensión. Las distribuciones o círculos gaussianos se pueden controlar utilizando lo que se denomina perplejidad, que afecta la varianza de la distribución (el tamaño del círculo) y el número de vecinos más cercanos, para el rango normal de perplejidad está entre el intervalo de 5 y 50 (Violante, 2018, p.7).

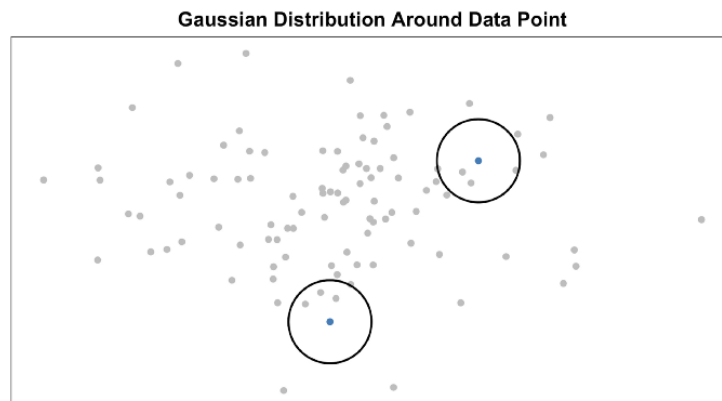


Figura 4-1: Medición de similitudes por pares en el espacio de alta dimensión

Fuente: Violante Andre, 2018.

Paso 2: El paso 2 es similar al paso 1, pero en lugar de utilizar una distribución gaussiana, se utiliza una distribución t de Student con un grado de libertad, también conocida como distribución de Cauchy (Figura 1-5). Esto nos da el segundo conjunto de probabilidades (Q_{ij}) en el espacio de baja dimensión. Como puede ver, la distribución t de Student tiene colas más pesadas que la distribución normal. La cola pesada permite una mejor simulación de largas distancias (Violante, 2018, p.7).

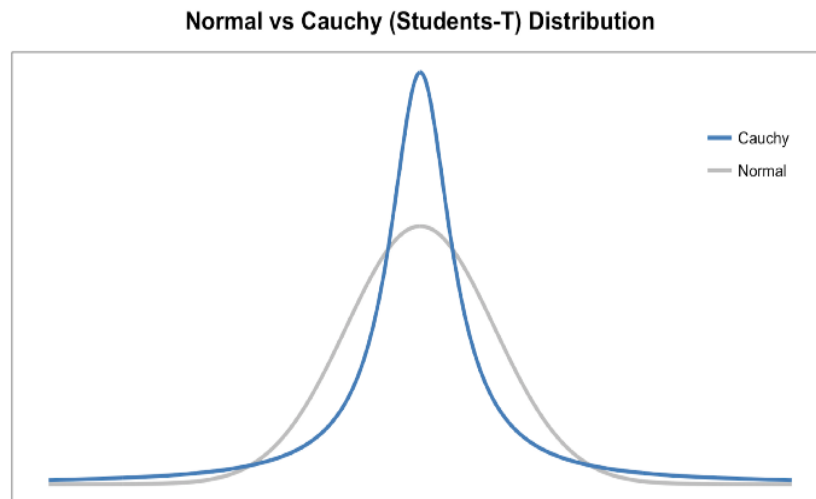


Figura 5-1: Distribución t normal vs. Student

Fuente: Violante Andre, 2018.

Paso 3: El paso final es pretender que el conjunto de probabilidades del espacio dimensional inferior (Q_{ij}) refleje lo más fielmente posible las probabilidades del espacio dimensional superior (P_{ij}). Se pretende que los dos mapas tengan una estructura similar. Se usa la divergencia de Kullback Liebler (KL) para medir la diferencia entre las distribuciones de probabilidad en un espacio bidimensional. El KL es un método asimétrico que compara de manera eficiente valores grandes de P_{ij} y Q_{ij} . Finalmente, se utiliza el descenso de gradiente para minimizar nuestra función de costo KL (Violante, 2018).

1.1.5.7. UMAP: Proyección y Aproximación de Múltiple Uniforme para Reducción de Dimensiones

Es un algoritmo de reducción de dimensionalidad no lineal que intenta preservar la estructura local de los datos. Tiene fundamentos teóricos basados en topología algebraica y geometría Riemanniana. Proporciona resultados competitivos con mejor cálculo y retención en comparación con el escalado t-SNE Estructura global para más datos (Bravo, 2020, p. 11). Una ventaja de UMAP es que tiene una base matemática que garantiza que una serie de propiedades topológicas relevantes que describen la variedad de puntos de alta dimensión se capturan de manera eficiente

mediante representaciones topológicas creadas para datos de baja dimensión (Cortez , et al., 2020, p. 7).

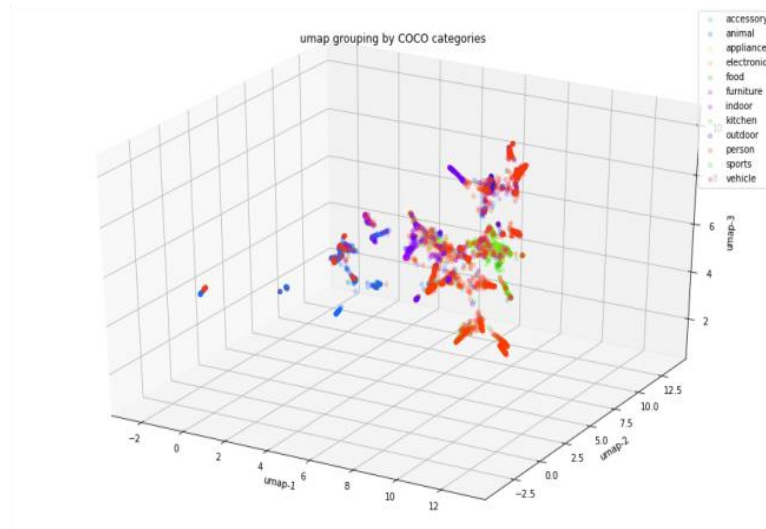


Figura 6-1: Proyección UMAP 3d

Fuente: ICHI.PRO, 2020.

Una técnica de reducción de la dimensionalidad que supone que las muestras de datos disponibles se distribuyen uniformemente en un espacio topológico (variedad), que se puede aproximar a partir de estas muestras de datos finitos y mapear (proyectar) a un espacio de menor dimensión (Dobilas, 2021). El secreto radica en la capacidad de UMAP para inferir estructuras locales y globales mientras preserva las distancias globales relativas en el espacio de menor dimensión. Estas habilidades permiten desbloquear conocimientos específicos, como encontrar similitudes entre las formas manuscritas de los dígitos 1 y 2 (Dobilas, 2021).

1.1.5.8. Base matemática de UMAP

El problema a resolver es encontrar una representación del conjunto de datos RN en el espacio dimensional inferior. Para hacer esto, asumimos que los datos son muestras de la variedad M de Riemman que luego se asignan al espacio de dimensión N (Bravo, 2020, p. 12).

- Los datos son una muestra estandarizada de M .
- M está conectado localmente.
- Existen suficientes datos que no hay un punto aislado, un punto está aislado si es el único punto de un componente M (M puede tener múltiples componentes no conectados) y tenemos una distancia tres en M que es localmente constante.

Dado que se quiere preservar la estructura de datos, se requiere aproximarlos de alguna manera a la distancia geodésica entre los puntos. Se tiene una distancia en el espacio mediano (RN) y

gracias a la premisa 4 también es una variedad de Riemann, es por eso que con estas distancias se logra acercar la distancia geodésica (Bravo, 2020, p. 12).

1.1.6. Técnicas Clustering

El Clustering consiste en la agrupación automática de datos. Es un tipo de aprendizaje automático no-supervisado. Consiste en la agrupación automática de datos. Al ser un aprendizaje no-supervisado, no hay una respuesta correcta. Esto hace que la evaluación de los grupos identificados sea un poco subjetiva (Heras, 2020).

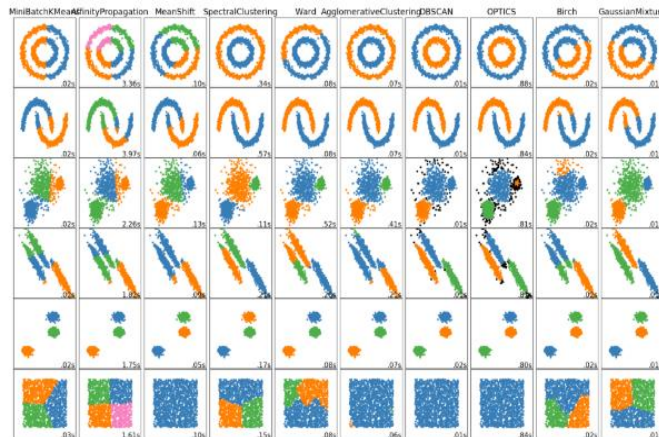


Figura 7-1: Técnicas Clustering

Fuente: Heras José, 2020.

También se conoce como análisis de conglomerados, tiene como objetivo agrupar elementos en grupos homogéneos en función de la similitud o similitud entre elementos. Las observaciones generalmente se agrupan, pero el análisis de conglomerados también se puede aplicar para agrupar variables. Estos métodos también se conocen como métodos de clasificación no supervisados o automáticos, también se conoce con reconocimiento de patrones no supervisados. El nombre no supervisado se usa para distinguirlos del análisis discriminante (Peña, 2002, p. 227).

1.1.7. Los principales algoritmos de Clustering

1.1.7.1. Diagrama de dispersión algoritmo UMAP agrupación K-Means

K-Means es uno de los algoritmos de aprendizaje no supervisado más simples que tiene como propósito ayudar a resolver un conocido problema de agrupamiento, este procedimiento sigue una forma muy simple y también fácil de clasificar un conjunto de datos, dado en algún conjunto de

prioridad fija (por ejemplo, k conjuntos), que se implementa en `sklearn.cluster.KMeans` (Kodinariya & Makwana, 2013, p. 91).

El propósito es definir los centroides k, uno para cada grupo, estos centros deben ubicarse inteligentemente porque diferentes ubicaciones conducen a diferentes resultados, por lo tanto, la mejor opción es ponerlo lo más lejos posible, después de esto es tomar cada punto perteneciente a un conjunto de datos en particular y relacionarlo con el punto central más cercano y si no hay más puntos destacados, se completa el primer paso y se realiza la predicción, también es necesario recalcular los k centros de los grupos resultantes de la etapa anterior. Después de estos nuevos k centroides, se debe crear una nueva correlación entre los mismos puntos de datos y el nuevo centroide más cercano. Se ha creado un bucle, también de este bucle, puede ver que los centroides k cambian de posición paso a paso hasta que no hay más cambios. Es decir, los centros ya no se mueven (Kodinariya & Makwana, 2013, p. 91).

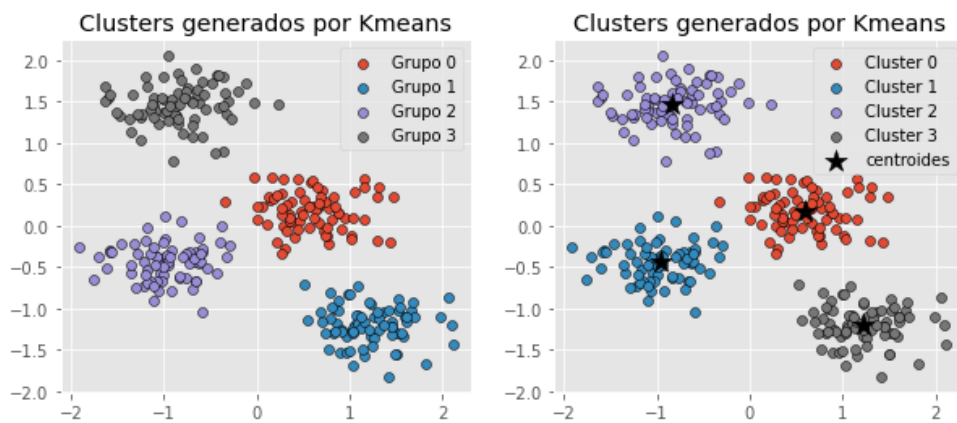


Figura 8-1: Agrupación K-means.

Fuente: Joaquín Rodríguez, 2016.

Este algoritmo pretende reducir una función objetivo, en este caso una función de error cuadrático. Su función objetivo (Kodinariya & Makwana, 2013, p. 91).

$$W(S, C) = \sum_{k=1}^K \sum_{i \in S_k} \|y_i - c_k\|^2 \quad (8-1)$$

En el cual S es una sección de los grupos K, también es del grupo de objetos representado por el vector y_i ($i \in I$) en el espacio de objetos de dimensión M, que consta de grupos S_k no vacíos y no repetitivos, cada uno con un centro c_k ($k = 1, 2, \dots, K$). El algoritmo incluye los siguientes pasos (Kodinariya & Makwana, 2013, p. 91)

- 1) En el k puntos en el espacio representado por los objetos agrupados. Estos puntos son el foco del grupo original (Kodinariya & Makwana, 2013).

- 2) Asigna cada objeto al grupo con el centro más cercano (Kodinariya & Makwana, 2013, p. 91).
- 3) Cuando todos los objetos se vean afectados, vuelva a calcular la posición de los k centros (Kodinariya & Makwana, 2013, p. 91).
- 4) Se repite los pasos 2 y 3 hasta que los ejes dejen de moverse (Kodinariya & Makwana, 2013, p. 91).

1.1.7.2. Diagrama de dispersión algoritmo UMAP agrupación DBSCAN

DBSCAN es un algoritmo de aprendizaje automático diseñado para detectar automáticamente 'clusters', elementos que están cerca unos de otros en función de sus propiedades en diferentes dimensiones, a diferencia de otros algoritmos de agrupamiento como K-Means, DBSCAN es adecuado para encontrar patrones de agrupamiento en el espacio físico. Un ejemplo es la distribución espacial de las actividades humanas. Como se muestra a continuación, entre las muchas alternativas, DBSCAN es la única que se aproxima a los resultados de la calificación de un analizador humano en el mapa (Vásquez, 2021, p.3).

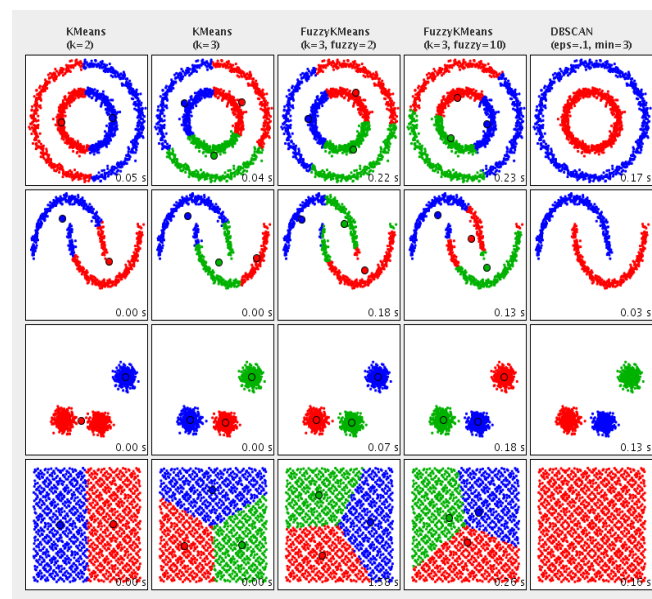


Figura 9-1: Agrupación DBSCAN

Fuente: Hipparchus, 2021.

1.1.7.3. El algoritmo DBSCAN se puede abstraer en los siguientes pasos

1. Encuentra todos los puntos cercanos en eps y selecciona el centro o los puntos visitados con más MinPts que vecinos.
2. Para cada concentrador, si aún no está asignado a un clúster, cree un nuevo clúster.
3. Localiza iterativamente todos sus puntos conectados a la densidad y los asigna al mismo grupo que el punto central. Se dice que los puntos A y B están densamente conectados si

el punto C tiene suficientes puntos en sus vecinos y ambos puntos A y B están dentro de una distancia eps. Es un proceso de secuencia. Entonces, si b es vecino de c, y c es vecino de d, y d es vecino de e, que a su vez es vecino de a, entonces b es vecino de a (Zhang, 2020).

4. Explora los puntos no visitados restantes en el conjunto de datos. Puntos que no pertenecen a ningún grupo ruidoso.

1.1.7.4. Modelo jerárquico distancia euclídea enlace Ward-Algoritmo UMAP

Este proceso de agregación tiene como objetivo crear grupos para que la suma de los cuadrados de las desviaciones de la media de cada variable (correspondiente en este caso a la estación de lluvia) sea el mínimo para todas las estaciones a la vez. Este procedimiento tiende a fusionar conglomerados con menos observaciones y formar conglomerados con aproximadamente el mismo número de conglomerados (Heredia, et al., 2012, p. 14).

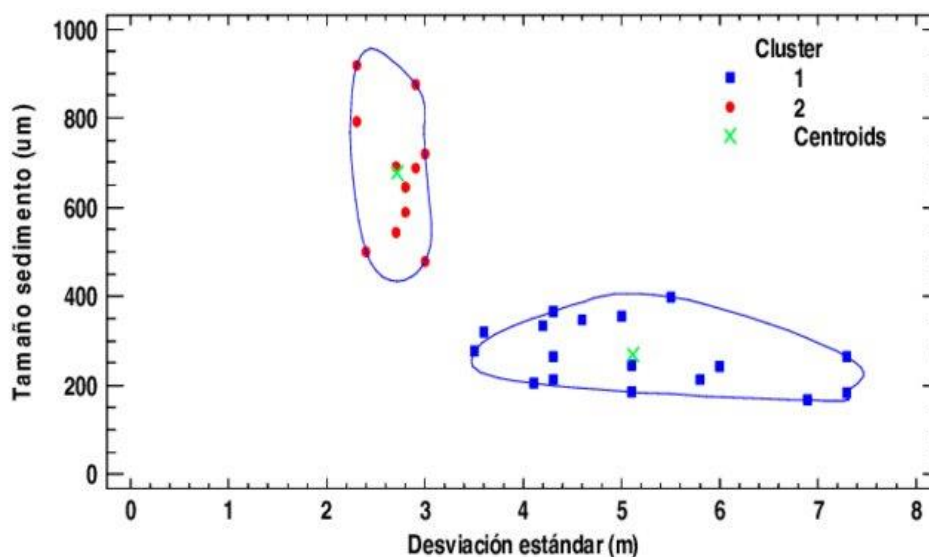


Figura 10-1: Análisis de conglomerados con el método de Ward y distancia Euclídea

Fuente: Almonacid Jaime, 2017.

Ward sugirió que la pérdida de información que ocurre cuando diferentes individuos se combinan en grupos se puede medir por la suma de los cuadrados de las desviaciones entre cada punto (individual) y la media de la masa en la que se integran. Para una agrupación óptima, en cada etapa del análisis se tiene en cuenta la probabilidad de unir cada par de grupos y se elige la unión de los dos grupos para aumentar al menos la suma de los cuadrados de las desviaciones en el proceso. El método de Ward es uno de los más utilizados en la práctica. Tiene casi todas las ventajas del método K-Means y generalmente es más discriminatorio al determinar los niveles de agregación. Una encuesta realizada por Kuiper y Fisher mostró que este método puede lograr una

clasificación óptima mejor que otros métodos (máximo, mínimo centro y promedio) (De la Fuente, 2011, p. 41).

Este mismo método lo que hace es reducir la distancia euclidiana al cuadrado del conjunto o grupo y esto quiere decir que también es un procedimiento jerárquico, todos los datos se agrupan en grupos con muchos elementos, la principal ventaja del método de Ward es: grupos más compactos de elementos similares, tamaño y miniatura, pérdida de información durante la organización del bloque, el cuadrado de la distancia euclidiana es la escala utilizada para formar la masa en este trabajo y se muestra (Heredia, et al., 2012):

$$d(i, j)^2 = \sum_k (x_{ik} - x_{jk})^2 \quad (9-1)$$

1.1.7.5. Optimizando UMAP- métrica chebyshev

La distancia de Chebyshev se delimita como la diferencia máxima entre dos vectores en cualquier dimensión de coordenadas. Es decir, es solo la distancia máxima a lo largo del eje, también es conocido como la distancia del tablero de ajedrez, porque el número mínimo de movimientos que necesita el rey para pasar de una casilla a otra es igual a la distancia denominada Chebyshev y la distancia de Chebyshev entre dos puntos n dimensionales se define como (Turing, 2020).

$$D_{cheb}(i, j) = \max_k |x_{ik} - x_{jk}| \quad (10-1)$$

Tabla 1-1: Ventajas y Desventajas de las Técnicas Clusters

Técnicas	Ventajas	Desventajas
Machine Learning		
K-Means	Fácil de entender y fácil de adaptar.	Se le pide al usuario que seleccione el número de clústeres.
DBSCAN	No necesita seleccionar el valor de clúster, además es robusta al momento de detectar outliers (observación anormal y extrema).	Utiliza la distancia euclidiana que no es efectiva porque llaman maldición de dimensionalidad, no es buen agrupado de datos.
Agrupación Aglomerativo	Deben indicar la distancia a emplear y el tipo de vinculación. Por eso los resultados varían.	Porque no deja ningún tipo de cabos sueltos, agrupa todos los clústeres.

UMAP-Métrica chebyshev	No mantiene completamente la densidad. Existe la posibilidad de crear desgarros inexistentes en grupos.	Para un mejor exploración y evaluación de agrupamientos.
---------------------------	---	--

Fuente: Google, 2022.

Realizado por: Miranda, Geovanny, 2022.

1.1.8. Instrumentos de procesamiento

1.1.8.1. Anaconda

Anaconda es una plataforma de ciencia de datos de código abierto que reúne las mejores herramientas para la ciencia de datos. Es una pila de ciencia de datos que incluye más de 100 paquetes populares basados en Python, Scala y R. Con la ayuda de su administrador de paquetes conda, los usuarios pueden trabajar con cientos de paquetes en diferentes idiomas y realizar preprocesamiento, modelado y agrupación de datos, fácil clasificación y verificación (Yan & Yan, 2018, p. 1).

1.1.8.2. Jupyter Notebook

Los cuadernos Jupyter, anteriormente conocidos como cuadernos Ipython, son un entorno informático interactivo que se puede utilizar para realizar análisis de ciencia de datos basados en Python con énfasis en la investigación reproducible. El entorno interactivo es excelente para el desarrollo y nos permite compartir cuadernos fácilmente (Sarkar, et al., 2017, p. 72).

1.1.8.3. Python

Este es un lenguaje de programación de alto nivel interpretado para programación de propósito general. Para análisis de negocios/ciencia de datos, Python es probablemente la primera opción entre muchos lenguajes informáticos emergentes (Yan & Yan, 2018, p. 2).

1.1.9. Modelo de Circulación Atmosférico

El modelo de Circulación Atmosférico manipulo datos de estaciones del INAMHI, datos proyectados CHIRPS, procesamiento de datos- formato lectura y análisis objetivo Creemas modificado para la interpolación de datos observados y post procesamiento de datos combinados en formato netCDF, Geotiff y TSV, donde los datos recolectados son en escala diaria del periodo

1985 a 2015 con 30 años, se disponen de 137 estaciones, tienen menos del 20% de datos faltantes para cada variable de análisis. Por lo tanto, nuestros datos fueron Tridimensional.

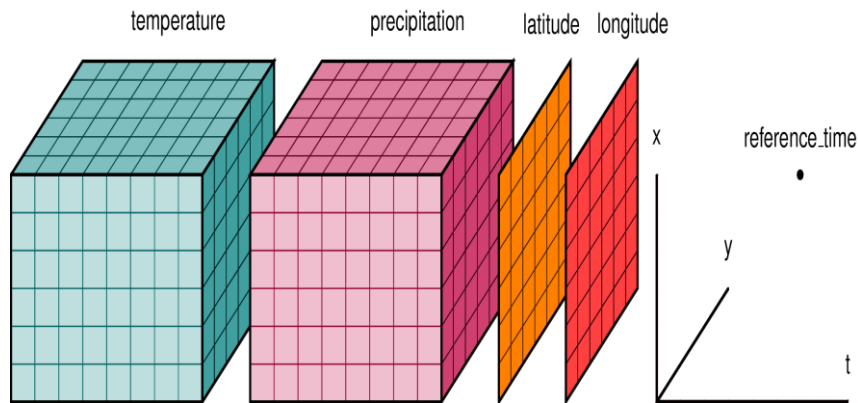


Figura 11-1: Datos Tridimensional

Fuente: Python, 2021.

1.1.9.1. Anomalías Climáticas

El autor utiliza un método Naive para calcular las desviaciones. Para entender las anomalías climáticas, primero debemos entender el comportamiento estacional Promedio. El siguiente gráfico muestra cómo cambia la temperatura para cada mes del año durante el período 1950 a 2000. Para ello se suman las temperaturas de todo el mes de enero de 1950 a 2000 y Promedios tomados de 25,5 °C (ver en la Figura 12-1) y así sucesivamente para los meses del año. Este promedio indica el ciclo anual promedio o comportamiento estacional para todo el año, por eso hay un valor medio de la temperatura referencial para cada mes del año (OMM, 2012).

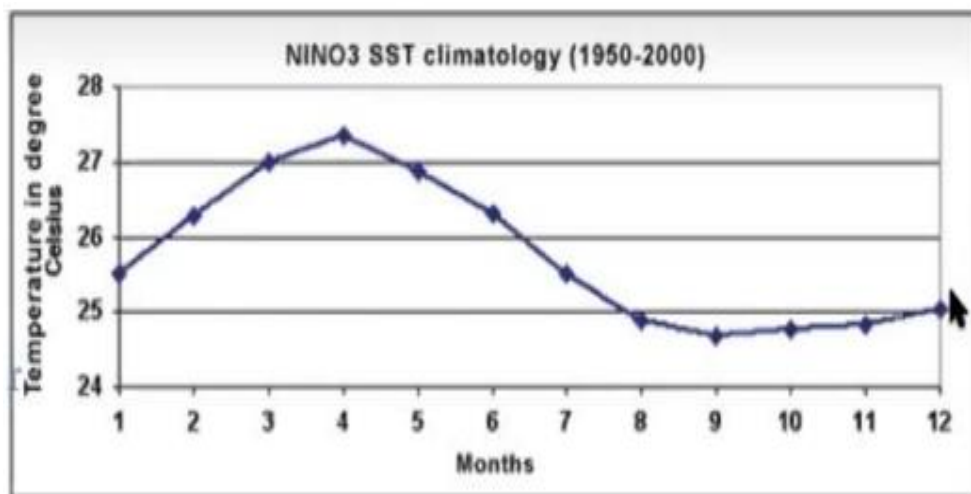


Figura 12-1: Anomalías de Climatología de 19950 a 2000

Fuente: Cnrm, 2014.

1.1.9.2. Anomalía

Si elige un año en particular y resta el comportamiento estacional promedio de ese año, obtiene una anomalía, como se puede observar en la Figura 13-1 existen anomalías positivas y negativas. En primer lugar, cabe señalar que las anomalías describen cómo esta variable se desvía del comportamiento estacional medio (OMM, 2012).

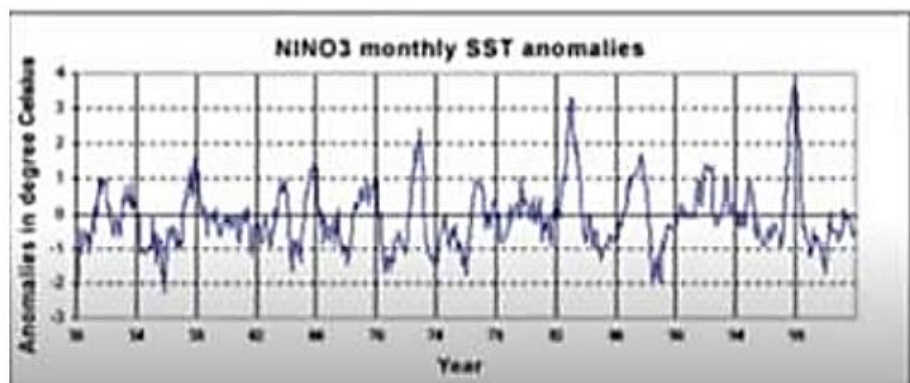


Figura 13-1: Anomalías positivas y negativas

Fuente: Cnrm, 2014.

1.1.9.3. ¿Por qué son útiles?

A lo largo de los años, meteorólogos y climatólogos de todo el mundo han creado y utilizado muchos indicadores de sequía, varían desde alguna medida simple, como el porcentaje de precipitación normal, los científicos en los Estados Unidos reconocen que los indicadores deben ser simples y fáciles de calcular, y también deben ser estadísticamente significativos y relevantes (OMM, 2012).

Cuadro 1. Valores del Índice normalizado de precipitación

2,0 y más	extremadamente húmedo
1,5 a 1,99	muy húmedo
1,0 a 1,49	moderadamente húmedo
-0,99 a 0,99	normal o aproximadamente normal
-1,0 a -1,49	moderadamente seco
-1,5 a -1,99	severamente seco
-2 y menos	extremadamente seco

Figura 14-1: Cuadro de valores del índice normalizado de precipitación.

Fuente: OMM, 2012.

1.2. Bases teóricas

1.2.1. Sistema Climático

El tiempo comúnmente se refiere al estado de la capa inferior de la atmósfera. (incluyendo el hábitat humano) en algún punto geográfico dado y en algún tiempo dado. Se describe en términos de los elementos meteorológicos directamente percibidos por los humanos, especialmente la temperatura del aire, la velocidad y dirección del viento, la cantidad y forma de las nubes, la cantidad y tipo de precipitación. A veces, la presión atmosférica, la humedad del aire y la también se citan el estado de la superficie terrestre (presencia de una capa de nieve, heladas, etc.), y en las regiones marítimas, la temperatura del agua y las fuerzas de las olas. El clima, por su parte, se refiere al conjunto de condiciones climáticas típicas de una región dada, junto con la frecuencia de estas condiciones y sus variaciones estacionales (Monin, 1986, p. 1).

1.2.2. Cambio Climático

El cambio climático y el calentamiento global y sus consecuencias están recibiendo una atención cada vez mayor a medida que aumenta la probabilidad de que ocurran y se predicen con mayor claridad las graves consecuencias del calentamiento global al nivel mundial (Vegara, 2010, p. 10), también se puede decir que es obvio, como lo demuestran las observaciones del aumento de la temperatura media global de la atmósfera y los océanos, y el papel predominante de la nieve y el hielo. El nivel del mar está subiendo en todos los mares de todo el planeta (Vegara, 2010, p. 17).

1.2.3. Dinámica Atmosférica

Dinámica atmosférica o también dinámica de la atmosfera, se conoce como la parte de la termodinámica, el estudio de las leyes físicas y el flujo de energía que intervienen en los procesos atmosféricos. Estos procesos son complejos debido a la gran cantidad de interacciones posibles entre la propia atmósfera y con el resto de nuestro planeta (sólido y líquido). La atmósfera está en capas, presentan características y propiedades específicas a diferentes niveles, cada capa es la convocatoria y es la siguiente: Estratosfera, troposfera, Mesosfera y Ionosfera (Pérez, 2006, p. 8). La presión es parte fundamental para dinámica atmosférica porque se define como la cantidad de fuerza que se ejerce sobre una unidad de área de alguna sustancia (Mott, 2006, p. 11). Por último, la densidad el aire también cumple con un rol importante, ya que es la relación entre la masa de una sustancia y su volumen y es una propiedad que suele expresarse en kilogramos por metro cúbico o gramos por centímetro cúbico, que varía más o menos según la presión, la temperatura y las condiciones (Etecé, 2021).



Figura 15-1: Dinámica Atmosférica

Fuente: Fernández Luis, 2006.

1.2.3.1. La humedad

La humedad está presente en la mayor parte de la atmósfera cerca de la superficie del planeta. No es algo que se pueda ver, pero se puede sentir indirectamente o incluso oler debido a la presencia del hongo (Tejada, et al., 2018, p. 33). Por lo tanto, cuando aumenta la humedad en el aire, se produce un movimiento vertical ascendente. Si la humedad disminuye, el aire se vuelve más denso y se produce un movimiento descendente. Estos movimientos debidos a cambios en la humedad se denominan convección de humedad. Para medir la humedad del aire se utilizan la humedad absoluta y la humedad relativa. La humedad absoluta son los gramos de vapor de agua en m^3 de aire [$HA = g \text{ H}_2\text{O (g)}/m^3$]. No tiene mucho valor porque la cantidad de vapor de agua que entra en el aire aumenta con la temperatura. Mientras que la humedad relativa de una masa de aire es

el porcentaje de vapor de agua, medido en correspondencia con el máximo que puede aguantar a esa temperatura (Picazo, 2021).

1.2.3.2. *Humedad relativa*

No especifica el contenido de vapor de agua por unidad de masa o volumen, simplemente expresa el porcentaje de vapor disuelto en un volumen de aire proporcional al vapor requerido para saturar el aire o expresa el cociente de presión de vapor y presión de vapor de saturación y se expresa como un porcentaje, un porcentaje más alto significa una mezcla de aire y agua más húmeda. Con una humedad relativa del 100%, el aire está saturado y en el punto de rocío (Tejada, et al., 2018, p. 36).

$$HR = 100 \frac{e}{e_s} \quad (11-1)$$

La velocidad del viento \vec{V} es un vector, donde la velocidad varía continuamente de forma aleatoria, tanto en módulo como en sentido y dirección. Tiene tres componentes según un triedro cartesiano, que se define según los siguientes criterios: los ejes x e y toman las direcciones E y N respectivamente, y el eje z toma la dirección cenital, solo se considera la componente de velocidad en la barra horizontal y se ignora la componente a lo largo del eje cenital (López, 2007, p. 42). La Temperatura está formada por átomos y moléculas en constante movimiento. El estado de la materia (sólido, líquido o gas) dependerá de las propiedades de las vibraciones moleculares, las sensaciones de calor o frío son quizás las más fáciles de representar del concepto macroscópico de temperatura (Petterssen, 1968, p. 40).

1.2.4. *La circulación general atmosférica*

La circulación general de la atmósfera es un sistema de vientos y presiones a escala planetaria debido a la variación de la radiación solar recibida, los efectos de la rotación de la Tierra y los efectos de la continentalidad. La circulación planetaria es en última instancia responsable del desarrollo y movimiento de los sistemas meteorológicos a pequeña escala (Nadal & Muñozuri, 2006, p. 105).

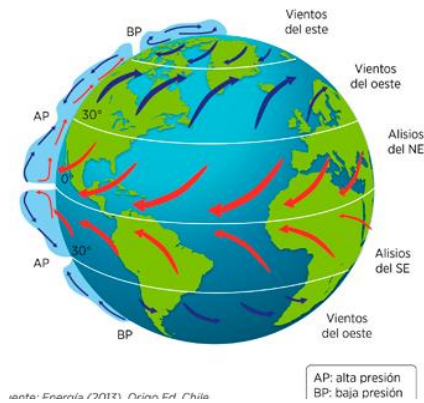


Figura 16-1: Circulación General de la Atmósfera.

Fuente: Energía, 2013.

1.2.5. Precipitación

Se emplea al término de precipitación para el agua líquida o sólida que llega al suelo desde la atmósfera. Así, son tanto la lluvia como la nieve, pero también el granizo, el rocío y la escarcha, aunque generalmente solo los dos primeros contribuyen significativamente a la precipitación total (Castillo & Jordán, 1999, p. 101).

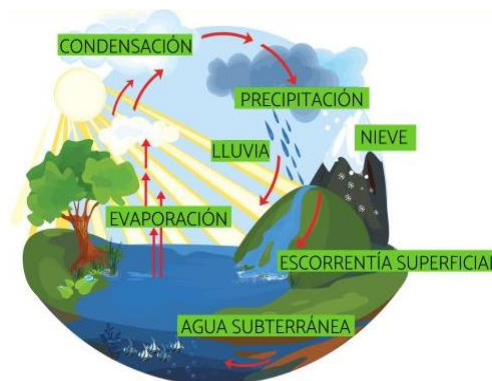


Figura 17-1: Precipitación

Fuente: Núñez Silvia, 2020.

1.2.5.1. Causas

- A medida que el aire de la superficie se calienta, su altitud aumenta. La temperatura de la troposfera disminuye con la altitud, es decir, cuanto más subimos, más fría se vuelve, por lo que a medida que la masa de aire asciende, se encuentra con el aire más frío y se satura (Portillo, 2017).
- Cuando las nubes que se desarrollan verticalmente se forman, las gotas de agua se adhieren a los núcleos de condensación, la masa de aire en la superficie no deja de ascender, ya que la cantidad de aire saturado y condensado finalmente aumenta en altura (Portillo, 2017).

- Cuando las gotas son lo suficientemente grandes como para vencer la resistencia, caen al suelo. Cuanto más tiempo permanecen las gotas en la nube, más grandes se vuelven a medida que se agrupan con otras gotas y núcleos de condensación (Portillo, 2017).

1.2.5.2. Mecanismo

Es el proceso de dar o transmitir un impulso inicial al pronóstico del tiempo para que comience un movimiento ascendente o convectivo, entendemos que el nombre que se le da a la precipitación es una especie según el mecanismo por el cual produce. Hay cuatro mecanismos principales por los cuales el viento se levanta y provoca la precipitación (Brenes & Saborío, 1995, p. 30).

- Convección térmica
- Convección orográfica
- Convección ciclónica
- Convección frontal

1.2.5.3. Latitud

El ángulo (norte-sur) entre dos líneas imaginarias, una de las cuales se extiende desde un lugar dado hacia el centro de la Tierra, y la otra se extiende desde el centro de la Tierra hacia el Ecuador (Butler, 1990, p. 24).

1.2.5.4. Longitud

El ángulo entre dos líneas imaginarias (este u oeste del meridiano de origen), una que se extiende hacia el eje de la Tierra y la otra hacia el meridiano de origen (MO) (Butler, 1990, p. 24).

1.2.5.5. Tiempo

Va desde unos pocos minutos hasta unas cuatro semanas, tiempo durante el cual la tormenta puede persistir, o la masa de aire puede permanecer estacionaria, o puede formarse un ciclón tropical, y en un momento dado, las condiciones atmosféricas pueden cambiar. Algunas estaciones y observatorios meteorológicas realizan mediciones meteorológicas por hora que pueden seguir la

evolución de la representación de los elementos atmosféricos, y los pronósticos a corto plazo, también el tiempo son provisionales (Contreras , 2001, p. 142).

1.2.6. Condensación

Es el proceso por el cual el vapor de agua de la atmósfera se convierte en agua líquida. La condensación no ocurre cuando la corriente de vapor de agua en la atmósfera alcanza el nivel de saturación, la condensación no se produce automáticamente. El vapor de agua necesita un material de soporte que pueda enfriar. Suelen ser las impurezas del aire las que facilitan la operación, y en ocasiones la condensación se forma directamente sobre la superficie del producto donde su temperatura está por debajo del punto de rocío. Algunas partículas de condensado son tan pequeñas que flotan en el aire, formando nubes, mientras que otras gotas de agua más grandes se precipitan en forma de lluvia, granizo o nieve (Aguilera, et al., 2020).

1.2.6.1. Coalescencia

Es la unidad fundamental de la precipitación a temperaturas extremas. En tiempo frío con temperaturas inferiores a -4°C , se forman pequeños cubitos de hielo con agua fría. La formación de cristales de hielo también requiere núcleos de condensación, pero los núcleos de condensación crecen más rápido porque la presión de vapor sobre los cristales de hielo es menor que la presión de vapor sobre el agua subenfriada. Los cristales crecen inmediatamente por colisión con agua helada subenfriada. Estos cristales forman copos de nieve que pueden llegar a lugares como la lluvia si la temperatura no es demasiado alta. La mayoría de las precipitaciones que ocurren en latitudes medias se basan en un proceso llamado teoría de Bergeron (Zúñiga & Crespo del Arco, 2021).

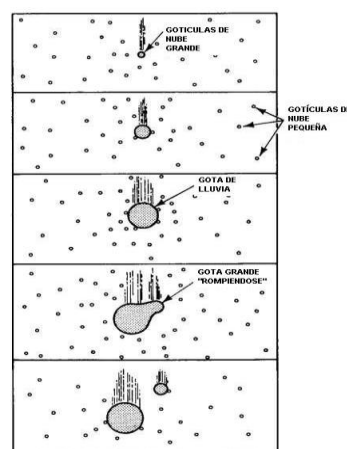


Figura 18-1: Coalescencia

Fuente: González Mónica, 2011.

1.2.7. Precipitación en el Ecuador

En las zonas costeras, la precipitación media anual varía desde menos de 60 mm (Península de Santa Elena) hasta más de 2000 mm dependiendo de la ubicación en el área afectada por el movimiento Humboldt, mientras que en la sierra el promedio anual de las precipitaciones varía entre 800 y 1500 mm, Con excepción del Valle del Chota (300 mm) y el Valle de Jubones (400 mm), las precipitaciones en la región amazónica son regulares, aumentando levemente entre marzo y julio y disminuyendo en agosto, en enero es el resultado de transiciones tropicales. La precipitación más alta del país (> 4500 mm) se registró en las regiones de Mera y Puyo, pertenecientes a la provincia de Pastaza. Las zonas donde recibe más precipitaciones son las faldas de las cordilleras. Las zonas más secas son los valles interandinos y las zonas costeras central y sur (Ron, 2018).

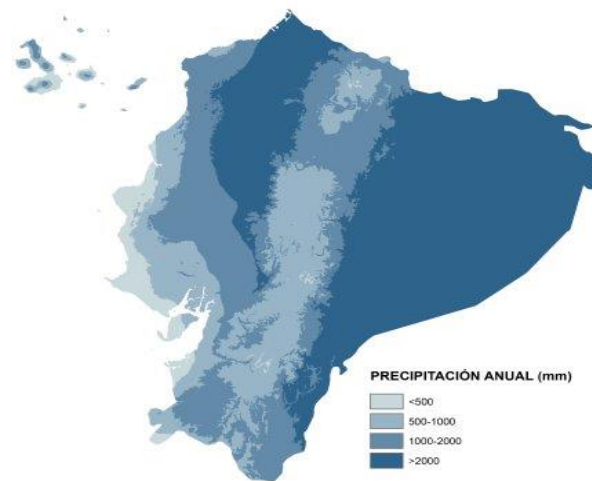


Figura 19-1: Precipitación Anual

Fuente: WorldClim, 2019.

En la noche del jueves 20 a la madrugada del viernes 21 de agosto de 2020, se presentó una fuerte inestabilidad atmosférica que provocó la formación de nubes sobre la Amazonía con fuerte crecimiento vertical, a diferencia de la troposfera, la temperatura es menor a 0 °C (INAMHI, 2020). La lluvia o precipitación en Ecuador pertenece a dos estaciones climáticas definidas invierno y verano, pero tienen una distribución específica para los meses del año; por lo tanto, un invierno severo en febrero a mayo e invierno templado en octubre hasta noviembre, mientras que un verano muy seco mientras que en junio hasta agosto y un verano templado conocido como veranillo del Niño durante el mes de diciembre hasta enero, esto se infiere al estudio de los antecedentes de lluvia correspondientes al período 1980 hasta 2006 proporcionados esta información por el INAMHI (Portilla, 2018, p. 79).

1.2.7.1. Humedad (relativa) en el Ecuador

La humedad relativa está íntimamente relacionada con la precipitación o disminución de temperatura como también de precipitación, de manera que los valores altos de humedad relativa corresponden a meses de invierno y valores bajos, corresponden a meses secos de término o verano. Un caso particular es en el Oriente, donde la humedad tiene un valor alto y constante como la estación M008 en Puyo situada a 920 metros de altitud logrando 90.7 % en junio y 87.11% en agosto, en todo el país la humedad relativa se mantiene por encima del 70% del promedio inferido. El promedio más bajo, pero aún por encima del 70% son estaciones de la región andina, como la estación M139 en Gualaceo, situada a 2.360 metros encima del nivel del mar, reportando 79.37% en abril y 74.97% en noviembre y se resume lo dicho sobre la correspondencia entre la altitud y la humedad relativa anual y confirma que en Ecuador no baja del 70% durante todo el año (Portilla, 2018, p. 103).

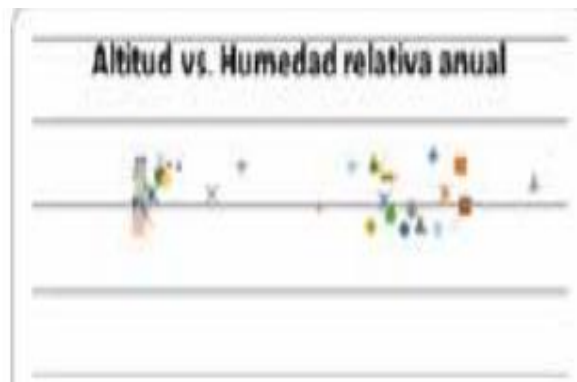


Figura 20-1: Altitud vs Humedad relativa Anual

Fuente: Portilla Fredi, 2009.

Típicamente, el porcentaje de humedad relativa es causado por la lluvia, y las estaciones en los Andes muestran más del 70%, pero no más del 85% de humedad, lo que afecta la presencia de otros factores y la proporción de humedad presente en el aire (Portilla, 2018, p. 103).

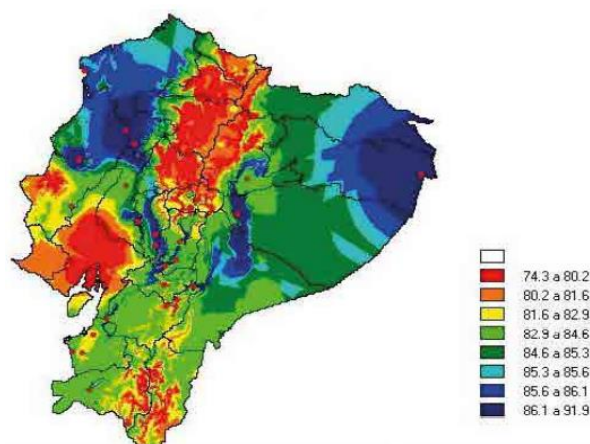


Figura 21-1: Humedad relativa anual-Ecuador

Fuente: Portilla Fredi, 2018.

En la (Fig. 21-1) muestra la distribución típica del porcentaje de humedad relativa en la región continental, los porcentajes más bajos son entre (0-10) se ubican en las regiones andinas norte y sur, y la costa sur del Ecuador como son Guayas, Santa Elena y Manabí. La región andina central, Chimborazo, Cañar y Azuay tiene un porcentaje mayor al 80% del promedio anual (30-50%). Lo mismo es cierto para las regiones centro-sur y este. En contraste, los valores más altos de humedad (86%-92%) se encuentran en las regiones centro nororiental y occidental de la Amazonía y la costa norte. Un punto a destacar es la presencia de humedad superior al 80% (30-50 por ciento) en la provincia del sur de Ecuador (provincias de El Oro y Loja), donde las precipitaciones son muy raras, con un promedio anual inferior a 1.000 mm (ver Figura 1.16), 20-30 por ciento); Así, por ejemplo, la estación M179 Arenillas ubicada en el sur a una altitud de 60 metros tiene una humedad anual del 82,5%.(Portilla, 2018, pp. 104-105).

1.2.8. Patrones climáticos que afectan el clima en el Ecuador (ENSO)

Los riesgos de Ecuador no solo se deben a los riesgos relacionados con su área geográfica en la línea media, sino también a sus malas condiciones económicas y culturales, su preparación para el próximo desastre y la importancia del cambio climático para los seres humanos (Toulkeridis, et al., 2020, p. 24). Durante el período 1960-2010, Ecuador reportó cambios en la ubicación y temporadas de precipitación como también de la temperatura todo el país, Con cada aumento del calentamiento global, el evento que ocurre una vez cada 100 años se convertirá en un evento anual hasta mediados de siglo en muchas regiones. Esto representa un riesgo para muchas provincias y las localidades de la costa baja están en riesgo (Telégrafo, 2019).

La "Corriente del Niño" se refiere a los afluentes del Océano Pacífico Sudamericano, moviéndose de norte a sur, hacia las costas como de Perú y así mismo del Ecuador cada año. Este fenómeno se considera beneficioso porque representa el tiempo húmedo como también el tiempo lluvioso,

que es importante para la agricultura. Los pescadores del sur de Ecuador y el norte de Perú mantuvieron este nombre del 18 al 25 de diciembre, cuando nació el niño Jesús (Paladines , et al., 2015, p. 101). En Ecuador distinguimos: aumento de eventos climáticos extremos, como el fenómeno del Sur llamado El Niño, acrecentamiento del nivel del mar, repliegue de los glaciares, mayor expansión del dengue y otros padecimientos tropicales, esparcimiento poblacional de variedades invasoras en Galápagos y otros ecosistemas sensitivos del Ecuador, y pérdida de especies (INAMHI, 2020).

Desde 1997 hasta junio de 1998 se registraron un total de 88.590 víctimas de El Niño y 6.153 familias gravemente heridas, que perdieron sus casas y pertenencias, y tuvieron que depender de albergues, familiares o amigos para su modo de vida. El fenómeno El Niño tuvo un impacto significativo en la población afectado en Ecuador (Paladines , et al., 2015, p. 102).



Figura 22-1: Corriente del Niño

Fuente: El Universo, 2020.

1.2.8.1. Anticiclón del pacífico

Cuando el Pacífico oriental está activo y cerca de tierra, los vientos alisios son más fuertes y el ZCIT es el más septentrional, está entre el 8°N y 10°N, durante los duros inviernos, los antioxidantes disminuyen, (Dobilas, 2021) los vientos alisios se debilitan y la ZCIT se acerca al ecuador. El Anticiclón Subtropical del Pacífico Sudeste (SPSA) se extiende a lo largo de la cuenca del Pacífico Sur y es un forzante importante del sistema de la Corriente de Humboldt. ASPSP presenta variabilidad estacional, interanual e Inter decadal (Inter decadal). Esta última variación está asociada con la Oscilación Decadal del Pacífico (PDO), considerada un patrón en la Cuenca del Pacífico (Ancapichún, et al., 2015, p. 1).

CAPITULO II

2. MARCO METODOLÓGICO

2.1. Tipo de la Investigación

La presente investigación es de tipo cuantitativa, que tienen un conjunto de procesos secuenciales como se observa en la Figura 1-2.

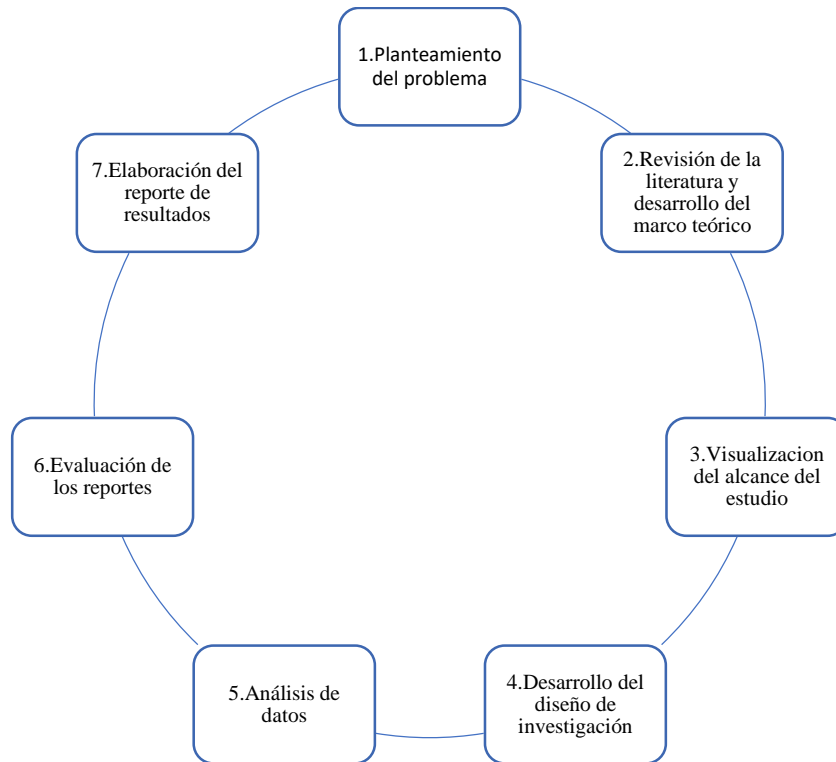


Figura 1-2: Proceso cuantitativo

Fuente: Hernández & Fernandez, 2016.

Realizado por: Miranda, Geovanny, 2021.

2.2. Planteamiento del problema

Una de los efectos del calentamiento global son las variaciones de la precipitación que pueden llegar a producir fuertes lluvias o sequías en algunas épocas del año, por este motivo es importante estudiar, descubrir y comprender los patrones meteorológicos del país.

2.3. Revisión de la literatura y desarrollo del marco teórico

Hoy en día no existen estudios que investiguen el comportamiento del clima en el Ecuador, se han encontrado pocos artículos científicos sobre técnicas machine learning que pronostican eventos. En el campo de la climatología, las técnicas de aprendizaje automático han sido muy

utilizadas, se han encontrado alrededor de mil artículos relevantes en la literatura científica, de los cuales se consideraron los 35 elementos, lo que facilitaría a contrastar los resultados obtenidos. Fuentes primarias consideradas, como 18 artículos científicos relacionados con el clima del mundo y el Ecuador; fuentes secundarias: se ha utilizado 5 libros de Python, 2 libros de análisis multivariante, 5 libros de cambio climático, 2 libros de método de investigación y fuentes terciarias: páginas web y videos tutoriales para la consulta códigos de lenguaje de programación.

2.4. Visualización del alcance del estudio

El alcance de este estudio es exploratorio en cuanto se propone descubrir patrones de comportamiento climático futuro en el Ecuador, y también es descriptivo de modo que se espera representar cómo se manifestarán en el tiempo los diferentes patrones de comportamiento encontrados.

2.5. Desarrollo del diseño de investigación

El diseño de investigación se define como no experimental pues busca comprender los patrones de comportamiento de la precipitación que se presentarán durante el período 2020-2050. Según el periodo temporal transversal en el cual se considera los años 2020-2050 para el análisis de las variables del fenómeno en estudio.

2.6. Análisis de datos

En cuanto al análisis de datos, se utilizó el lenguaje de programación Python, para la elaboración histogramas de frecuencia, gráficos de líneas, se efectuó técnicas multivariadas para la reducción de dimensionalidad, PCA, TSNE y UMAP con métricas (Chebyshev y Euclídea). Asimismo, se aplicó diferentes técnicas Machine Learning: K-Means, DBSCAN y agrupación Aglomerativo para observar el comportamiento de los patrones de precipitación. Finalmente, el modelo se validó utilizando métrica de precisión y gráficas.

2.7. Evaluación de los reportes

La fiabilidad del algoritmo ha sido calculada y evaluada mediante métricas de precisión como: Silhouette y el método de codo para determinar el número óptimo de clústeres.

2.8. Elaboración del reporte de resultados

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

2.9. Diseño de la investigación

Se respalda en un diseño no experimental la investigación es cuantitativa-exploratorio, porque se trabaja con observaciones climáticas y la ubicación de estaciones meteorológicas a través de predicciones, patrones de comportamiento, se basan en información cuantificable, considerando la información proporcionada por investigaciones anteriores, esto genera un modelo que puede estudiar el comportamiento de los patrones de precipitación durante los últimos 30 años (Berger et al. 2018).

2.10. Localización de estudio

El Ecuador se encuentra en América del Sur, por lo que su territorio se distribuye en dos hemisferios. Está restringida al norte con Colombia, al sur y al este con Perú y al oeste con Océano Pacífico.



Figura 2-2: Mapa del Ecuador

Fuente: Google maps, 2022.

2.11. Población de estudio

La población del estudio son anomalías de la precipitación de todo el Ecuador del periodo 2020-2050.

2.12. Método de muestreo

La información con la que se va a trabajar es adquirida directamente del modelo de circulación atmosférico CMIP6, razón por la cual no se aplica un método de muestreo.

2.13. Tamaño de la muestra

Se realiza el estudio con los datos del periodo 2020 a 2050 del modelo de circulación atmosféricos CMIP6, por lo tanto, no se toma una muestra.

2.14. Técnica de recolección de datos

Los datos utilizados por el modelo de circulación atmosférica CMIP6 son obtenidos de la biblioteca de datos UCSB CHIRPS y datos del Inamhi.

2.15. Identificación de variables

Variable dependiente

- Precipitación

Variables independientes

- Ubicación geográfica
- Tiempo

2.16. Modelo estadístico

Los modelos estadísticos que se han utilizado en la investigación: exploración de datos para conocer las características generales de la variable precipitación, la selección de técnicas multivariantes para reducción de la dimensión, aplicación de técnicas machine learning como el conglomerado para identificar los patrones de comportamiento de la precipitación, y la validación del modelo mediante métricas estadísticas y graficas.

2.17. Variables en estudio

2.18. Operacionalización de variables

Tabla 1-2: Operacionalización de variables

Nombre de la variable	Descripción	Tipo de variable	Escala de medición	Categoría o intervalo
Precipitación	Datos atmosféricos	Cuantitativa	Intervalo	$[-\infty, +\infty[$
Ubicación geográfica	Coordenadas geográficas	Cuantitativa	Intervalo	$[-\infty, +\infty[$
Tiempo	Periodos de tiempo	Cuantitativa	Intervalo	$+\infty[$

Elaborado por: Miranda, Geovanny, 2021.

CAPÍTULO III

3. MARCO DE RESULTADOS Y DISCUSIÓN DE RESULTADOS

A partir de los datos de la variable precipitación. Se muestra a continuación, exploración de los datos, determinar áreas o patrones para mejor toma de decisiones. Luego, se ha emplea técnicas de reducción de dimensión para la variable en estudio. Después, se aplicó técnicas machine learning para identificar patrones de comportamiento. Finalmente, la validación del modelo con métricas estadísticas y graficas.

3.1. Análisis Descriptivo

3.1.1. Variables cuantitativas

Tabla 1-3: Tabla de contingencia de la variable Precipitación Tridimensional

Dimensiones (T: 10954, X: 34, Y: 22)					
Coordenadas:					
*Y	(Y) float32	-16.11	-11.91	9.105	13.31
*X	(X) float32	-10.5	-101.2	-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: Miranda, Geovanny, 2022.					

La variable precipitación posee tres dimensiones diferentes con sus respectivas coordenadas donde: Y es latitud, X es longitud y T es tiempo. Además, los datos son valores que fueron recolectadas en escala diaria.

3.1.2. Descripción de datos

La base de datos de precipitación proviene de los modelos atmosféricos del Centro Nacional de Investigaciones Meteorológicas (CNRM) que se puede encontrar en el proyecto CMIP6, son datos de precipitación del periodo 2020-2050, de formato NetCDF el cual almacena datos multidimensionales como las coordenadas: latitud, longitud y tiempo. La variable precipitación

posee tres dimensiones, es decir es multidimensional, los datos fueron recolectados a escala diaria de 137 estaciones meteorológicas del INAMHI, que tienen menos del 20% de datos faltantes, alcanzando 30 años de datos.

3.1.3. *Estadística Descriptiva*

Para poder trabajar con una gama tan amplia de matrices Bidimensional, se tiene que obtener el promedio ya sea por fila o por columna para intentar describir lo que sucederá.

Tabla 2-3: Estadística Descriptiva

Estadístico	Resultado
Media (\bar{x})	-0.000083
Desviación estándar (s)	1.075627
Mínimo	-2.734937
Cuartil 25%	-0.711899
Cuartil 50% (mediana)	-0.190785
Cuartil 75%	0.486054
Máximo	6.902043

Realizado por: Miranda, Geovanny, 2022.

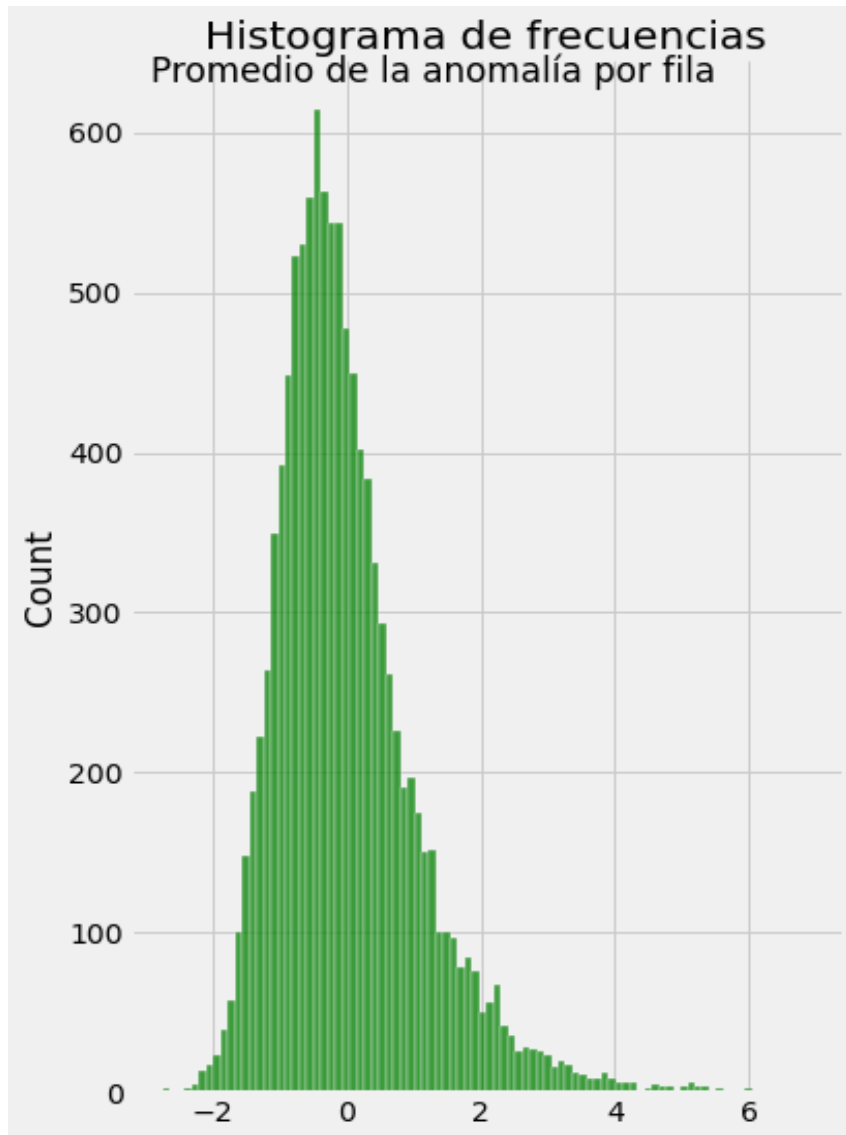


Gráfico 1-3: Histograma de frecuencias de promedio de anomalías por filas

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 1-3, el histograma de frecuencias de promedio de anomalías se corrobora que la distribución de las probabilidades del conjunto de datos es Normal Estándar.

3.1.4. Exploración de las anomalías

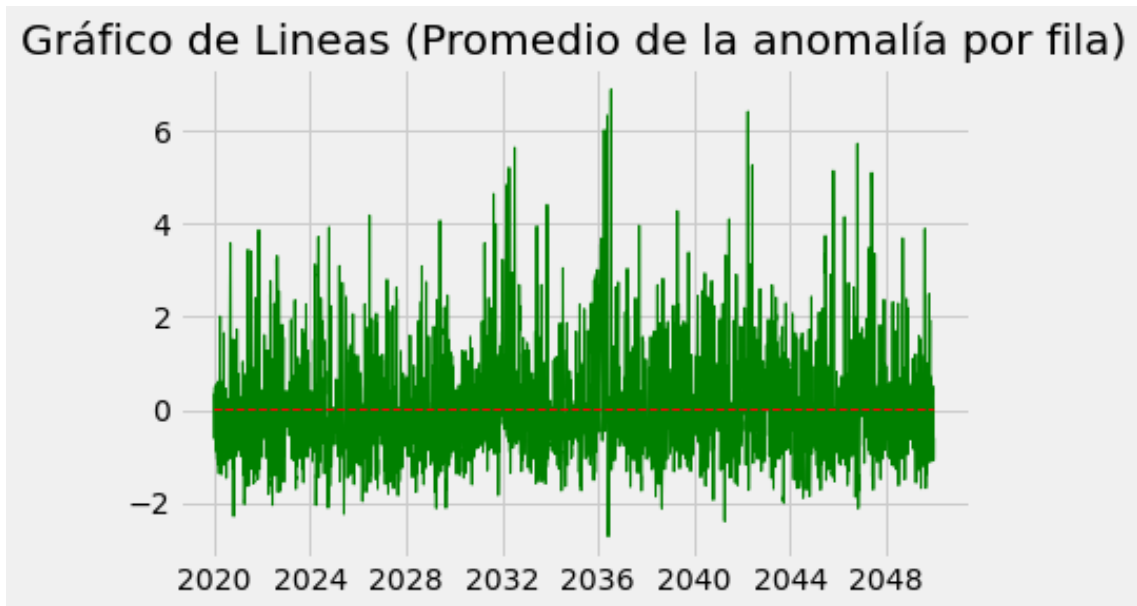


Gráfico 2-3: Gráfico de líneas promedio de anomalías por año

Realizado por: Miranda, Geovanny. 2022.

En el Gráfico 2-3, se visualiza días lluviosos (precipitación fuerte) en los picos en los años 2036-2040 como también sequías (extremadamente seco) desde el año 2020-2050. También se puede observar que el rango de las variaciones positivas(inundaciones) es más grande que las variaciones negativas(sequias).

Se agregan etiquetas de fecha a cada uno de los datos, luego se procedió a calcular los promedios mensuales en la base de datos, y finalmente se apila todas las variables y se traza el gráfico.

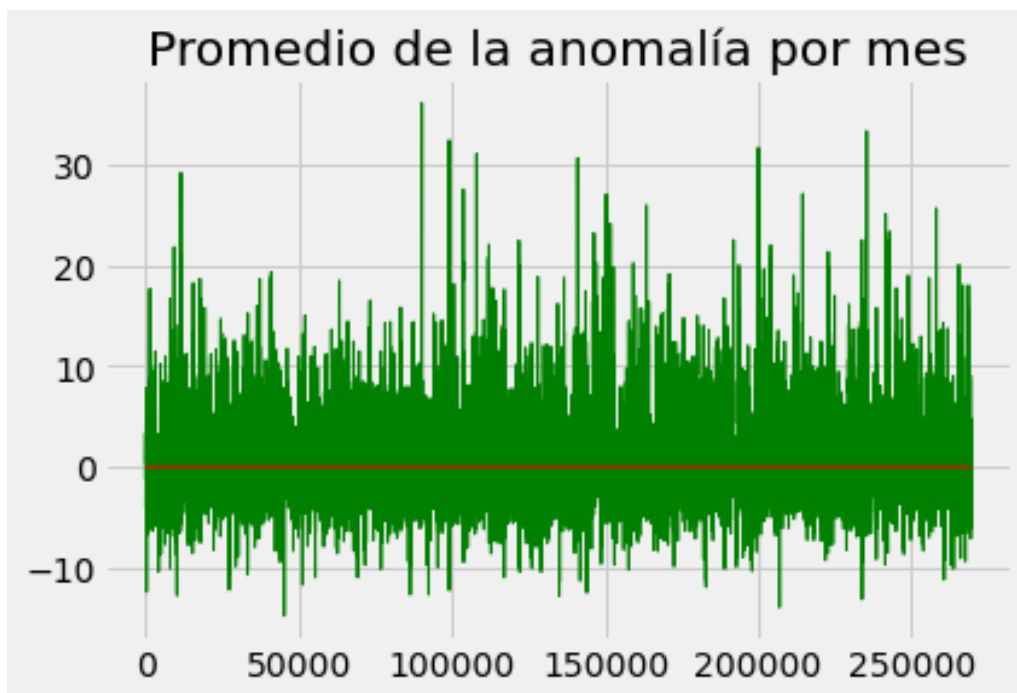


Gráfico 3-3: Promedio de anomalías por mes.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 3-3, se observa el comportamiento de las precipitaciones en donde se corrobora que existe meses con tendencia a fuertes lluvias y otros con tendencia a fuerte sequías.

3.1.5. *Matriz bidimensional*

Tabla 3-3: Dimensiones de la matriz Bidimensional de Precipitación

	Total, de filas	Total, de Columnas
Adif	10954	748

Realizado por: Miranda, Geovanny, 2022.

Para un buen análisis de la información se procede a convertir la matriz tridimensional a bidimensional, para una mejor interpretación y creación de un buen modelo Clustering se debe reducir la dimensionalidad de los datos, teniendo en cuenta que posee 10954 filas con 748 columnas, lo ideal sería obtener la mayor representatividad de la información.

3.1.6. *Reducción de dimensionalidad*

Una vez realizado la exploración de datos se evidencio que la matriz de datos es tridimensional, y se procedió aplicar la técnica de componentes principales para la reducir la dimensión.

3.1.7. Componentes principales

Para esta técnica es necesario determinar el porcentaje de varianza a explicar con las primeras componentes, en este caso se eligió el 90%.

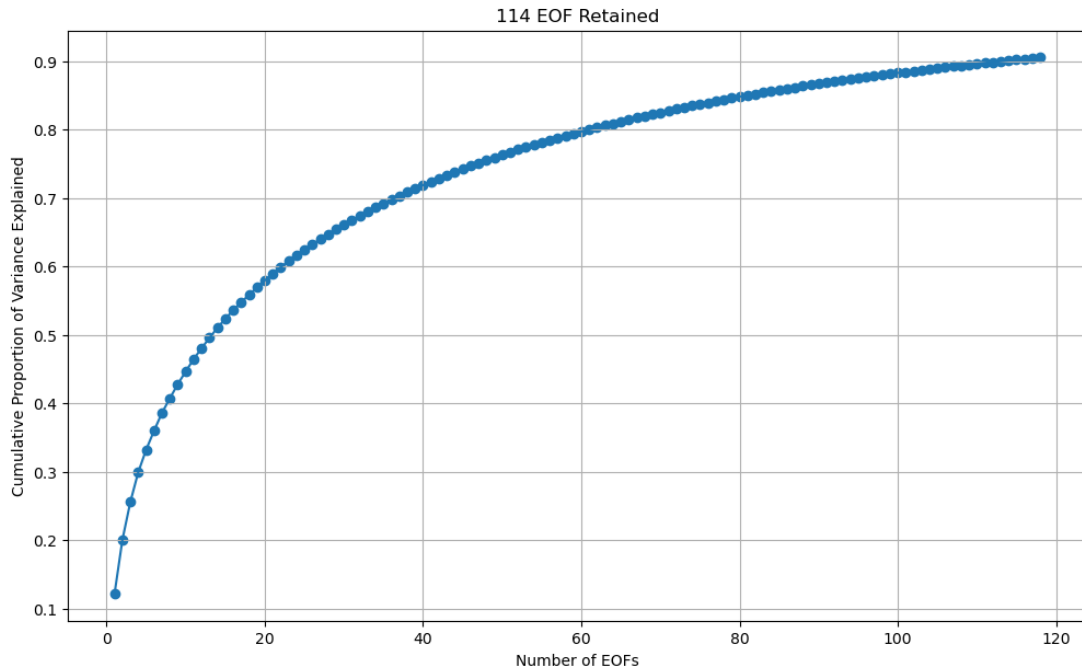


Gráfico 4-3: Componentes Principales.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 4-3, las tres componentes principales explicaron menos del 50% de variabilidad, desde el componente 5 en adelante se vuelven intrascendentes. Se necesitan aproximadamente 114 componentes para retener el 90% de la varianza. Se evidencio que los datos son inherentemente (no se puede separar) de alta dimensión: no se pueden describir de forma lineal con pocos componentes.

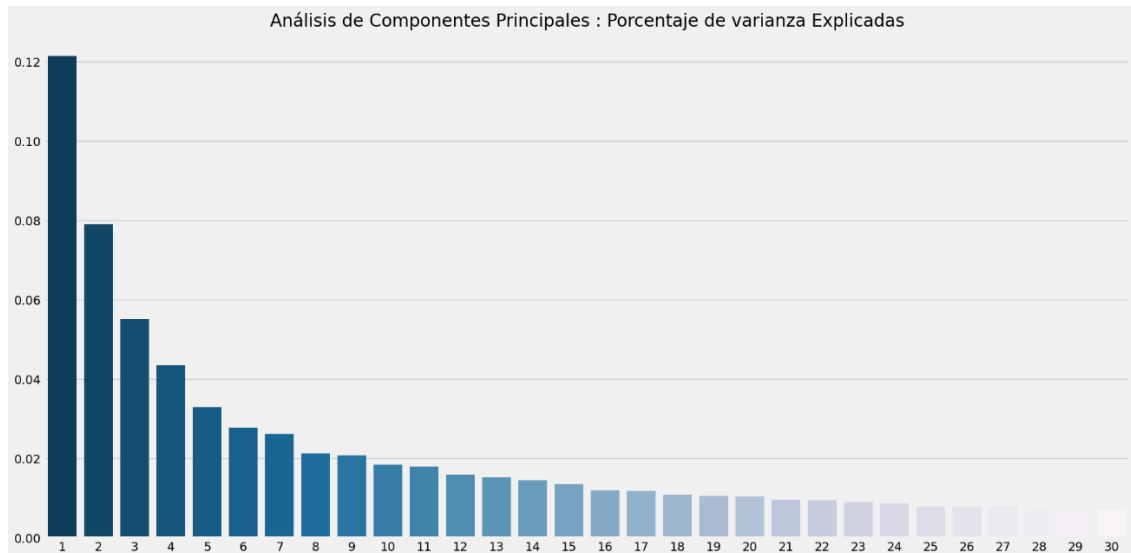


Gráfico 5-3: Porcentaje de varianza Explicada.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 5-3, se observa que la representación de las 2 primeras componentes principales explica aproximadamente el 20.045 % de la variabilidad.

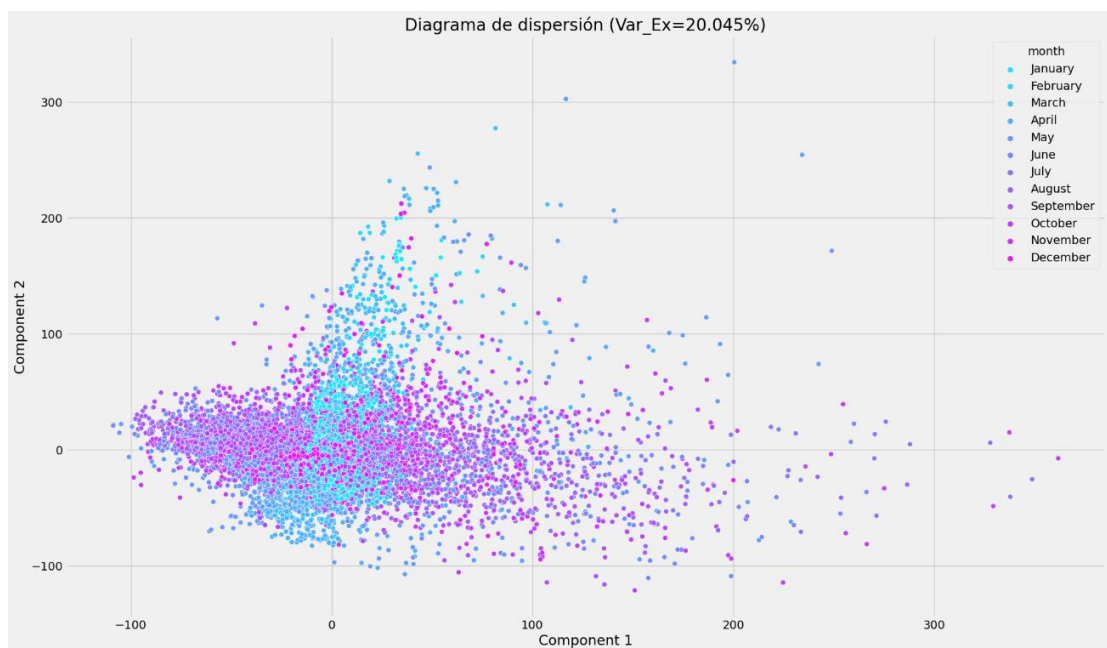


Gráfico 6-3: Diagrama de dispersión de los 2 primeros componentes principales por año.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 6-3, se observa el comportamiento uniforme de la precipitación, pero algunas de las fechas se desvían del comportamiento normal y el año no contribuye mucho en estas dos dimensiones, no existe agrupación de patrones.

3.1.8. *Técnica de Incrustación de vecinos estocásticos distribuidos en t (t-SNE)*

Es una técnica no supervisada utilizada principalmente para la exploración, visualización y reducción de datos de alta dimensión.

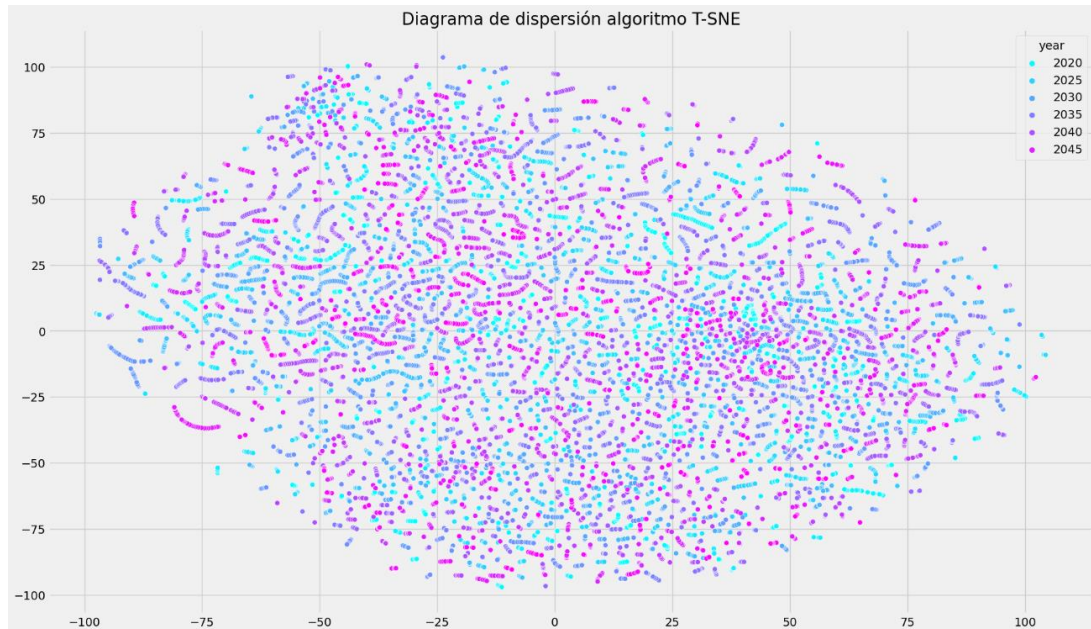


Gráfico 7-3: Técnica de Incrustación de vecinos estocásticos t-SNE por año.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 7-3, se observa que existe similitud entre los primeros años y los años posteriores, pero no existe agrupamiento de patrones, esta técnica se enfoca principalmente en la exploración y visualización de datos.

3.1.9. *UMAP: Proyección y Aproximación de Múltiple Uniforme*

Esta técnica permite interpretar mejor los datos y su estructura. Es un algoritmo para reducir conjuntos de datos de más de 3 dimensiones.

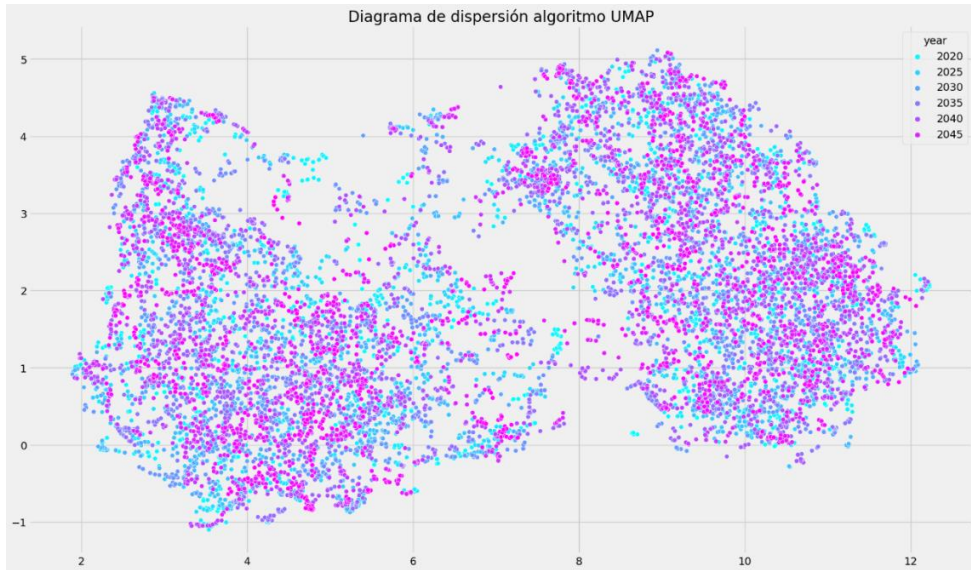


Gráfico 8-3: Diagrama de dispersión algoritmo UMAP por años.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 8-3, se observa que la técnica UMAP, ha identificado 6 patrones de comportamiento de la precipitación que se puede explicar como 6 tipos de climas.

3.1.10. Aplicación de Algoritmos de Clustering

Los parámetros se han optimizado con UMAP y los resultados se expresan con DBSCAN, que es lo más prometedor, e idealmente deberían ajustarse a las etiquetas de año y mes para comprender mejor el comportamiento de los patrones.

3.1.11. Selección del número óptimo de Clusters



Gráfico 9-3: Gráfica de numero de clústers.

Realizado por: Miranda, Geovanny, 2022.

En la Gráfica 14-3, se representa en el eje de la abscisa el número de clusters y en el eje de las ordenadas la suma de las distancias, el método del codo nos indica el número óptimo de clusters que para este estudio sería de 4.

3.2. Algoritmo UMAP (métrica=euclídean) agrupación DBSCAN

Se obtiene el siguiente gráfico por grupo.

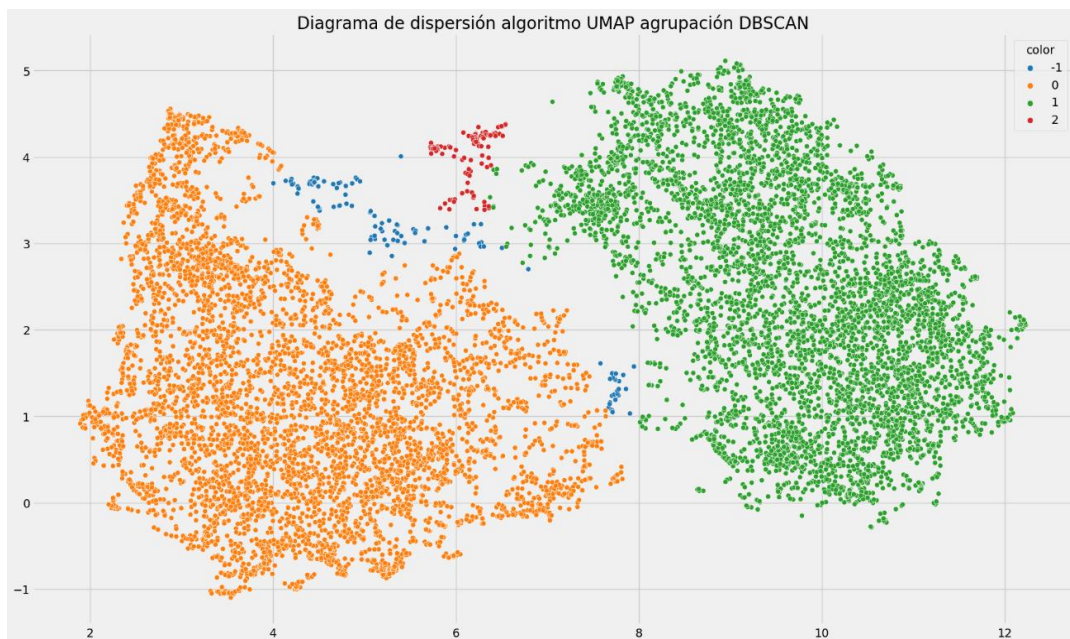


Gráfico 10-3: Diagrama de dispersión UMAP agrupación DBSCAN

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 10-3, se observa la dispersión de los patrones de precipitación, existe cuatro diferentes agrupamientos de puntos o patrones de precipitación, claramente se observa que existe poca distancia entre clusters donde indica: el valor -1 de color azul representa clúster 1 mientras que el valor 0 de color naranja es clúster 2, el valor 1 de color verde es el clúster 3 y el valor 2 de color rojo es el clúster 4.

Se utilizó métrica chebyshev este método es ideal para la agrupación de datos, se etiqueta los años por rango y se obtiene la siguiente tabla:

Tabla 4-3: Rango de años por cada grupo

Grupos	Rango de años	Cantidad de observaciones
-1	2020-2025	3481
0	2026-2035	3204
1	2036-2040	2594
2	2041-2049	1675

Realizado por: Miranda, Geovanny, 2022.

3.2.1. Matriz de confusión

Tabla 5-3: Matriz de confusión

		Valores Reales			
		2020-2025	2026-2035	2036-2040	2041-2049
Valores predichos	2020-2025	4016	8	304	12
	2026-2035	19	3653	225	42
	2036-2040	41	26	1826	7
	2041-2049	14	340	35	1070
Accuracy		0.8534			

Realizado por: Miranda, Geovanny, 2022.

El valor 0.85 de precisión indica que este algoritmo DBSCAN es un buen estimador, para identificar los patrones de comportamiento de la precipitación que han sido clasificados en cuatro grupos, dado que se tiene un enfoque de aprendizaje no supervisado, la matriz de Confusión se utilizó simplemente es para observar la predicción del algoritmo DBSCAN.

3.2.2. Representación de resultados por días del algoritmo DBSCAN

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

Grupos	Rango de años	Cantidad de observaciones predichos	Frecuencia relativa	Patrones
-1	2020-2025	3481	32.0%	Moderadamente Seco
0	2026-2035	3204	29.0%	Normal Aproximadamente Normal
1	2036-2040	2594	24.0%	Muy Húmedo Moderadamente Húmedo

2	2041-2049	1675	15%	Extremadamente Húmedo
---	-----------	------	-----	--------------------------

Realizado por: Miranda, Geovanny, 2022.

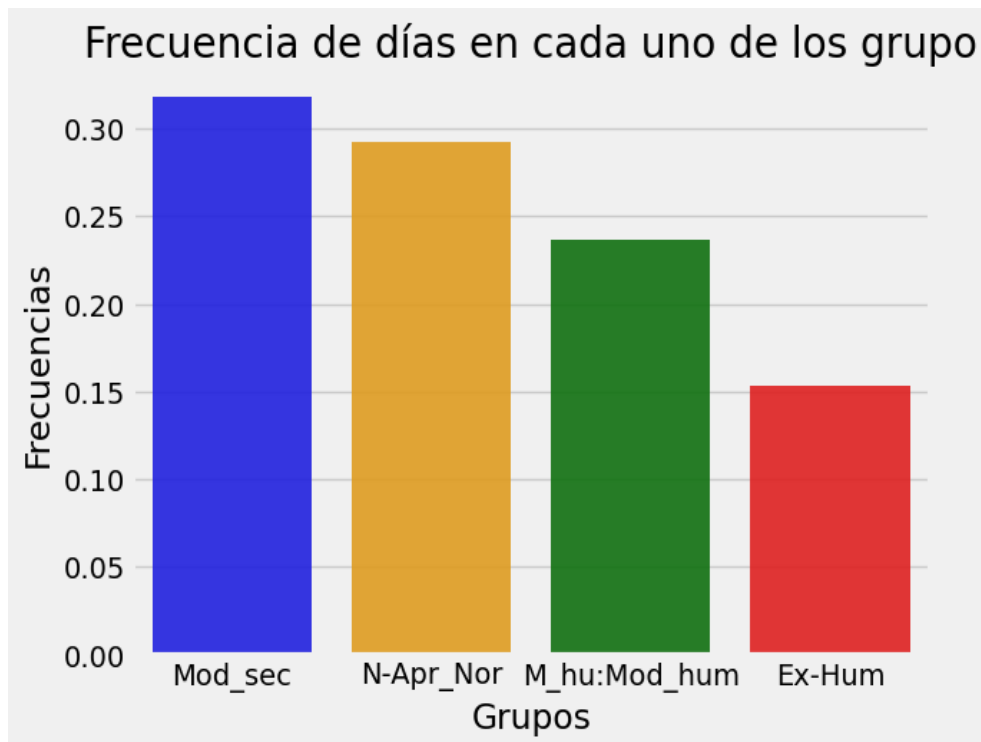


Gráfico 11-3: Frecuencia de días en cada uno de los grupos

Realizado por: Miranda, Geovanny, 2022

- Los días de sequías (moderadamente seco) forma el grupo -1(2020-2025), ya que por lo general Ecuador no tiene días extremadamente seco.
- La precipitación normal o aproximadamente normal (lluvias menos frecuentes) forma en el grupo 0 (2026-2035).
- La precipitación muy húmeda (lluvias frecuentes) forman el grupo 1(2036-2040).
- La precipitación extremadamente húmeda (lluvias intensas) forman el grupo 2 (2041-2049).

3.2.3. Mapa de calor de patrones de comportamiento

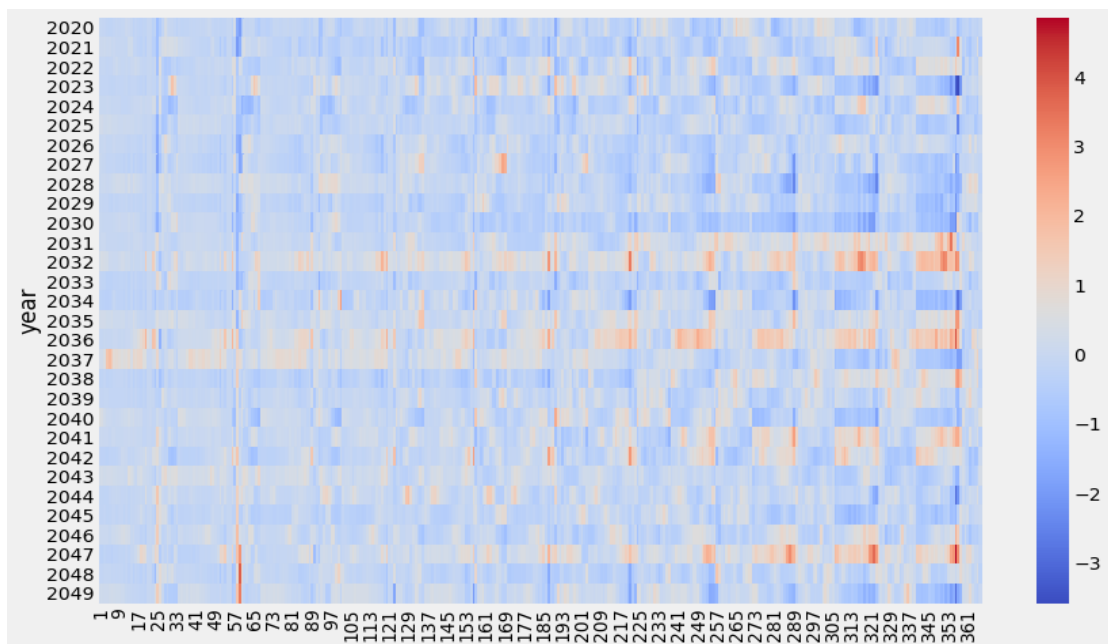


Gráfico 12-3: Mapa de calor de Patrones de comportamiento.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 12-3, se observa que en el año 2044 día 353 hay escasos días de sequías (extremadamente seco), mientras que en el año 2032 día 353 se visualiza días lluviosos (extremadamente húmedo). En el año 2030 existe un aumento de sequías en comparación con el resto de años.

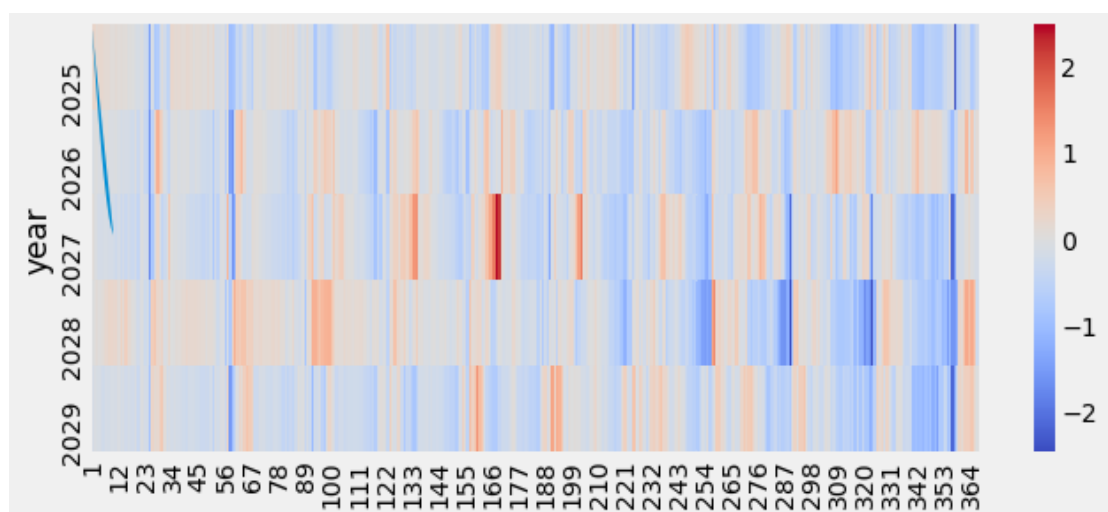


Gráfico 13-3: Mapa de calor Patrones de comportamiento.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 13-3, en el año 2027 día 166 se visualizó lluvias menos frecuentes (moderadamente húmedo), mientras que en el año 2028 día 320 se observó escaso día de sequía (moderadamente seco), mientras que en el año 2029 aumento de sequias y lluvias frecuentes en comparación al resto del año.

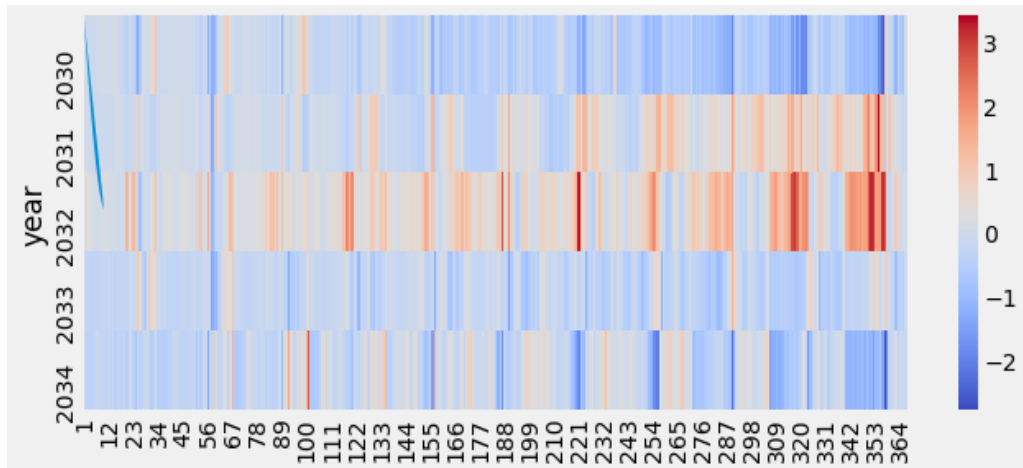


Gráfico 14-3: Mapa de calor Patrones de comportamiento.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 14-3, en el año 2030 día 56 se visualizó escasos días de sequía (moderadamente seco), mientras que en el año 2034 día 133 se observó pocos días de lluvia (moderadamente húmedo) mientras que en el año 2032 existirá lluvias frecuentes que sequias en comparación al resto de años.

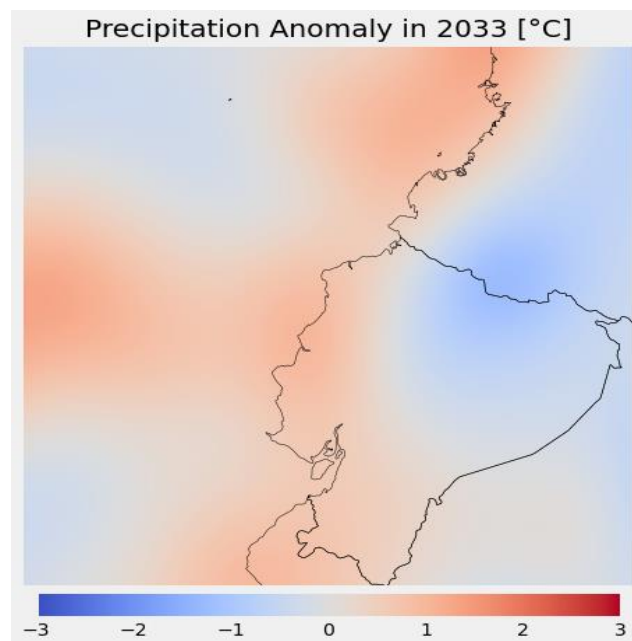


Gráfico 15-3: Mapa de calor Patrones de comportamiento.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 15-3, en la región costa en el año 2033 se visualiza lluvias frecuentes (muy húmedo), mientras que en las regiones de sierra y oriente se observa días de sequías (moderadamente seco).

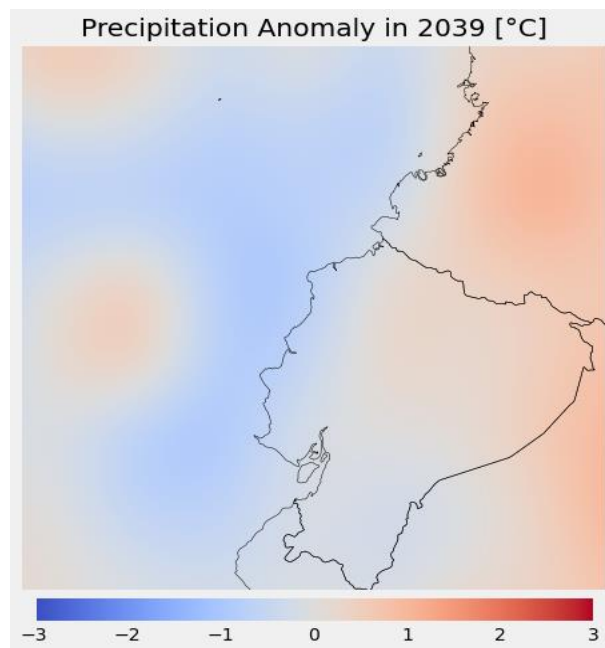


Gráfico 16-3: Mapa de calor Patrones de comportamiento.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 16-3, en el año 2039 en la región costa y una parte de la sierra se visualiza días de sequías (moderadamente seco), mientras que en la región sierra y oriente se observó menos frecuencia de lluvias (muy húmedo).

3.2.4. Algoritmo UMAP (métrica -euclídean) agrupación K-Means

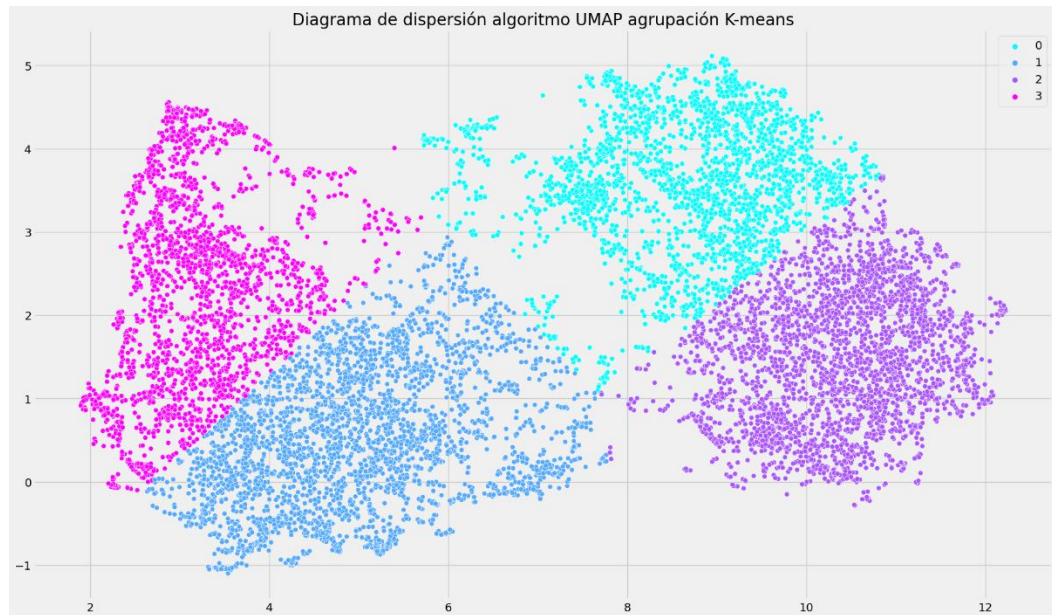


Gráfico 17-3: Diagrama de dispersión algoritmo UMAP agrupación K-Means.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 17-3, la dispersión de los datos representa la existencia de tres agrupamientos de puntos o patrones de precipitación, se evidencia que no existen un buen agrupamiento.

3.2.5. Algoritmo UMAP (métrica euclídean) agrupación Agglomerative Clustering

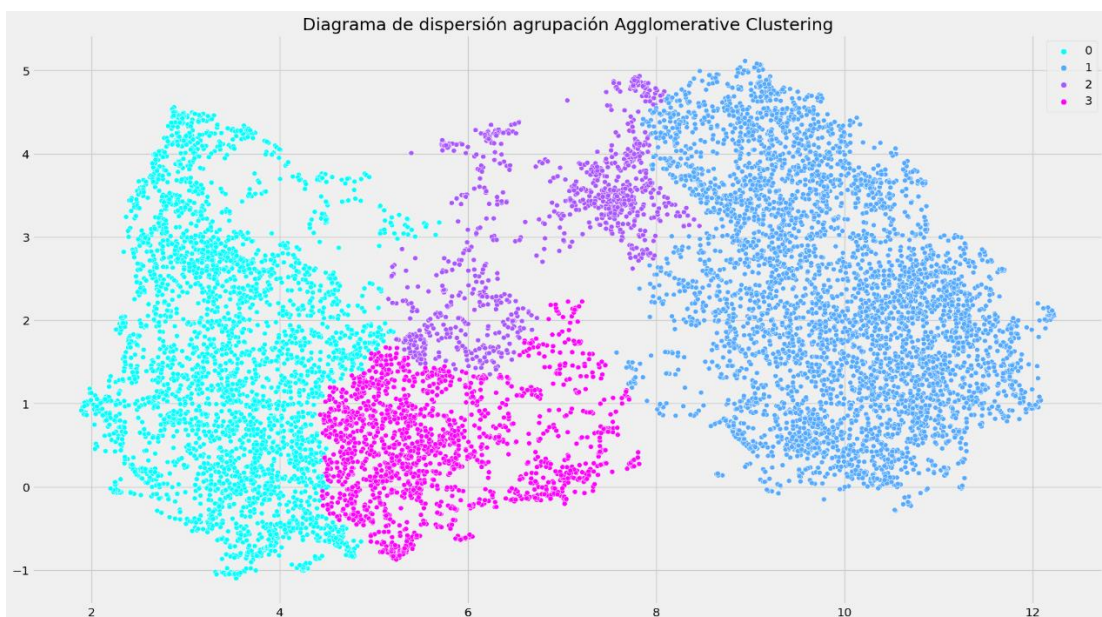


Gráfico 18-3: Diagrama de dispersión algoritmo UMAP agrupación Agglomerative Clustering.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 18-3, se visualiza las características de las observaciones en 4 grupos definidos.

3.2.6. Algoritmo UMAP (métrica-chebyshev) agrupación -K-Means

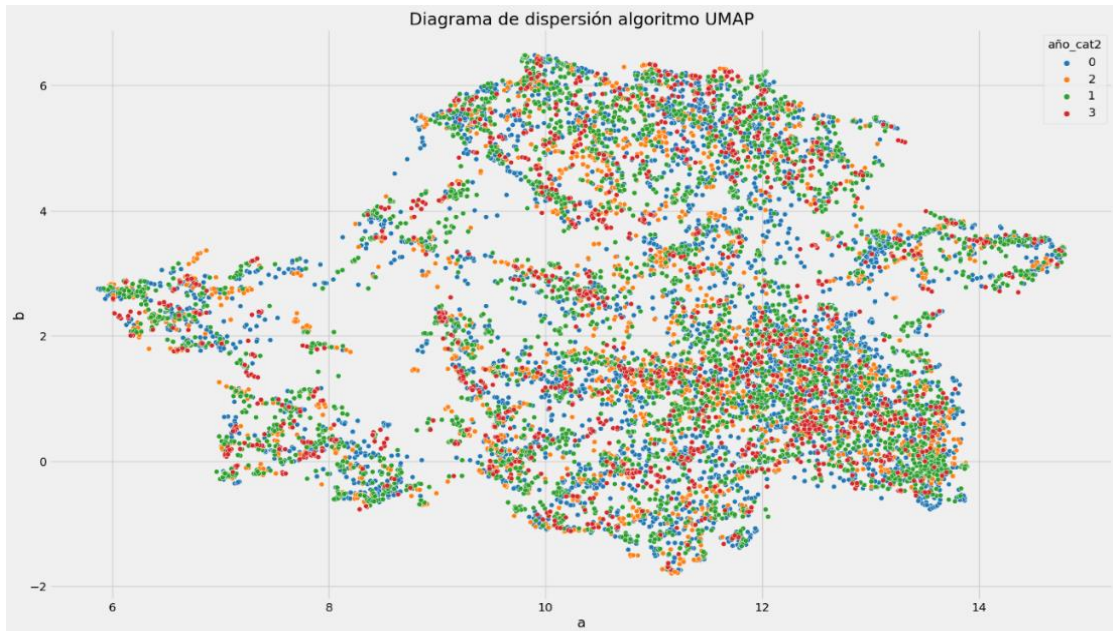


Gráfico 19-3: Diagrama de dispersión UMAP agrupación K-Means por categoría.

Realizado por: Miranda, Geovanny, 2022.

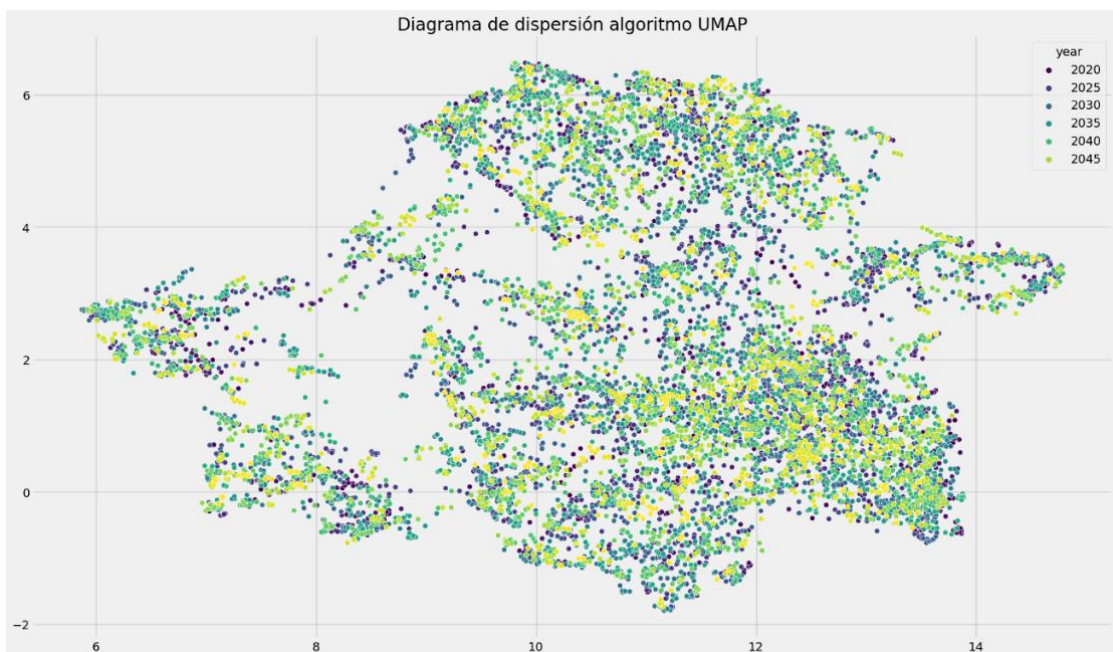


Gráfico 20-3: Diagrama de dispersión UMAP agrupación K-Means por años.

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 19-3 y Gráfico 20-3, se observa que no existe diferencia de clúster, ni separación de la información en grupos, por lo tanto, no se evidencio distancia entre los clústeres de los patrones de comportamiento de la variable precipitación por categoría y año.

3.2.7. Algoritmo UMAP (métrica-chebyshev) agrupación Agglomerative Clustering

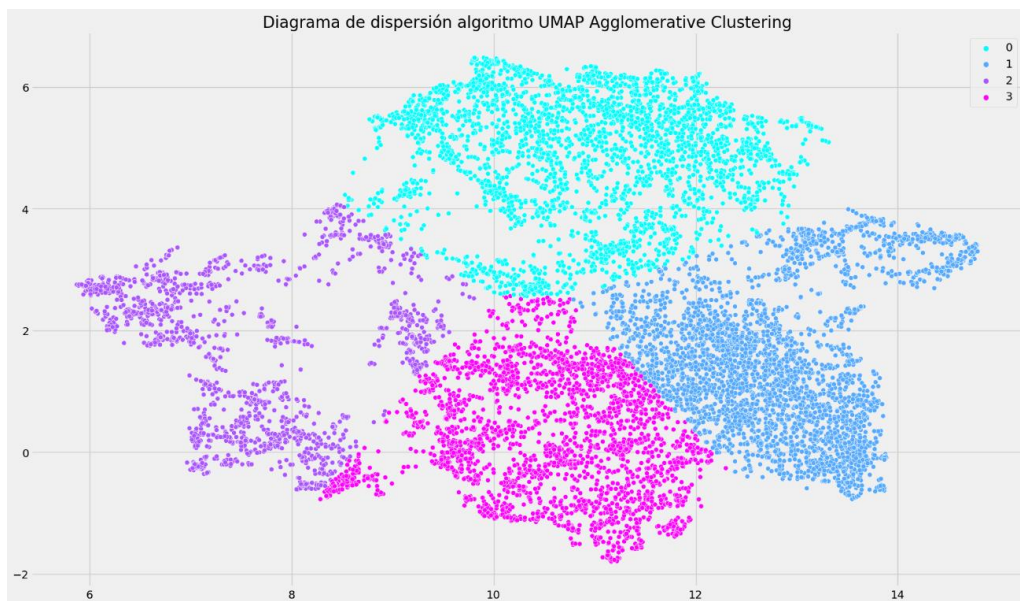


Gráfico 21-3: Diagrama de dispersión UMAP agrupación Agglomerative Clustering

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 21-3, se emplea la métrica chebyshev, se puede observar que existe diferencia de clusters, inadecuada separación de la información en grupos, por lo tanto, no se evidencia distancia entre los clústeres de los patrones de comportamiento de la variable precipitación.

3.2.8. Algoritmo UMAP (métrica-euclídea) agrupación DBSCAN

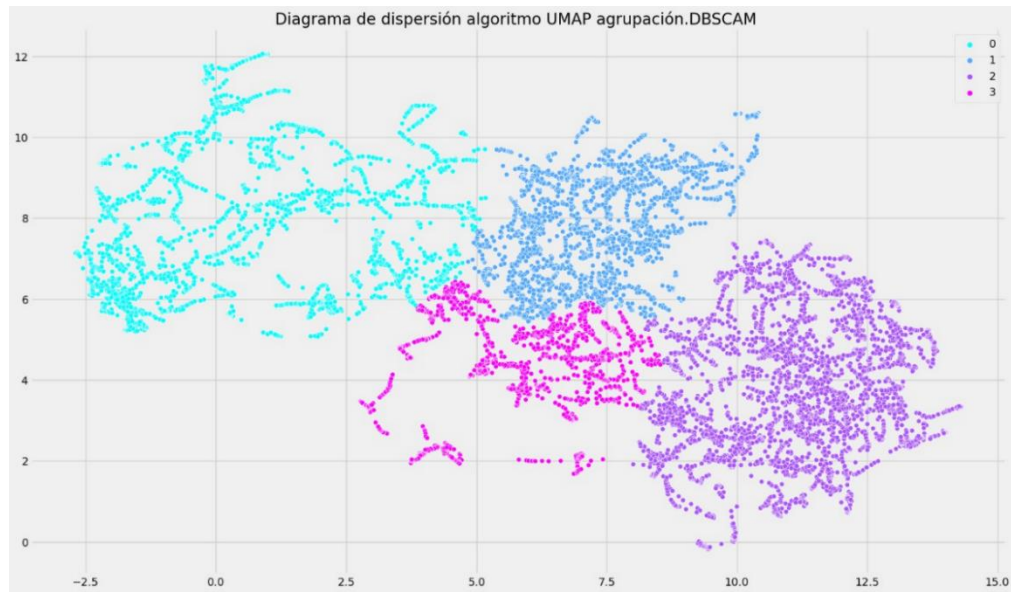


Gráfico 22-3: Diagrama de dispersión UMAP agrupación DBSCAN

Realizado por: Miranda, Geovanny, 2022.

En el Gráfico 22-3, se emplea la métrica chebyshev, se visualiza la inadecuada separación de la información en grupos, por lo tanto, no se evidencia distancia entre los clústeres de los patrones de comportamiento de la variable precipitación.

3.2.9. Validación de Modelo

Tabla 7-3: Índice de silhouette

Características UMAP	Algoritmo	Grupos	Silhouette
UMAP metric=" euclídea"	DBSCAN	4	0.7092
UMAP metric=" euclídea"	Agglomerative Clustering	4	0.4106
UMAP metric="euclídea"	k-means	4	0.4676
UMAP metric=" chebyshev"	k-means	4	0.5128
UMAP metric=" chebyshev"	Agglomerative Clustering	4	0.4383
UMAP metric="chebyshev"	DBSCAN	4	0.3690

Realizado por: Miranda, Geovanny, 2022.

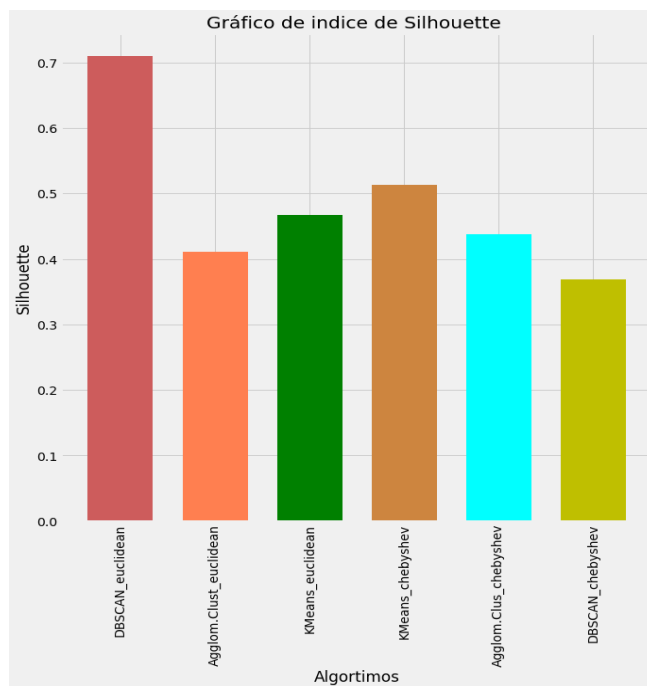


Gráfico 23-3: Índice de silhouette

Realizado por: Miranda, Geovanny, 2022.

Como se visualiza en la Tabla 7-3, los diferentes índices de Silhouette que fueron calculados para diferentes algoritmos de agrupación, la mejor técnica de algoritmo de agrupación es UMAP-métrica chebyshev- DBSCAM con una efectividad de 70.97%, por lo que se selecciona este algoritmo para observar el comportamiento de patrones de precipitación.

Tabla 8-3: Precipitaciones registradas a nivel nacional mes de enero y febrero de 2021

Región	Precipitación
Litoral y perfil costero	Bajo la normal, sobre la normal y normal
Interandina	Bajo la normal, Sobre la normal y normal
Amazónica	Bajo la normal y Sobre la normal.
Insular (San Cristóbal)	Bajo la normal y normal

Fuente: INOCAR, 2021

Realizado por: Miranda, Geovanny, 2022.

Tabla 9-3: Precipitaciones registradas a nivel nacional marzo - abril de 2021

Región	Precipitación
Litoral y perfil costero	Dentro y sobre la normal
Interandina	Dentro y sobre la normal, especialmente en norte de la región
Amazónica	normal.
Insular	normal.

Fuente: INOCAR, 2021

Realizado por: Miranda, Geovanny, 2022.

Tabla 10-3: Precipitaciones registradas a nivel nacional mayo- agosto de 2021

Región	Precipitación
Litoral y perfil costero	Normal y Bajo normal
Interandina	Normal, Sobre la Normal y Bajo La Normal.
Amazónica	normal.
Insular (San Cristóbal)	normal.

Fuente: INOCAR, 2021

Realizado por: Miranda, Geovanny, 2022.

Tabla 11-3: Precipitaciones registradas a nivel nacional septiembre- diciembre de 2021

Región	Precipitación
Litoral y perfil costero	Bajo la normal y normal, Sobre la Normal.
Interandina	Normal, Sobre la Normal.
Amazónica	Normal y Bajo la normal.
Insular	Normal

Fuente: INOCAR, 2021

Realizado por: Miranda, Geovanny, 2022.

La información de precipitaciones registradas en el Ecuador del año 2021 a nivel nacional se obtuvo de INNOCAR (www.inocar.mil.ec). Efectivamente, los comportamientos de precipitación son similares a los pronósticos de patrones de comportamiento realizado por la técnica Machine Learning DBSCAN, donde: el grupo -1 pertenece a días de sequía (moderadamente seco), grupo 0 lluvias menos frecuente (normal o aproximadamente normal), grupo 1 lluvias frecuentes (moderadamente húmedo o muy húmedo), por último grupo 2 lluvias intensas (extremadamente húmedo), se puede afirmar que el Algoritmo DBSCAN es la mejor técnica para predecir los patrones de comportamiento de precipitación. Por lo tanto, en Ecuador la precipitación tendrá una variación significativa provocando inundaciones y sequías. Existe dos estaciones climáticas que es el invierno y el verano, pero existe una distribución en común en los meses del año, como en febrero-marzo se registran un invierno intenso, mientras que en los meses de octubre-noviembre un invierno leve, en los meses de junio-agosto un verano muy seco y el fenómeno del niño se presenta en los meses de diciembre y enero.

CONCLUSIONES

- Se realizó una exploración de datos en el cual se describe la matriz de información: formato y estructura, se efectuó el algoritmo de reducción de la dimensionalidad UMAP métrica euclídea como la mejor opción.
- Se identificó 4 patrones de comportamiento mediante la técnica DBSCAN, estos grupos podría pertenecer a la frecuencia de días con precipitación: Moderadamente Seco (sequias), Normal-Aproximadamente Normal (lluvia poco frecuente), Muy Húmedo o Moderadamente Húmedo (lluvias muy frecuentes) y Extremadamente Húmedo (lluvias frecuentes).
- Se concluye que la mejor técnica para identificar patrones de comportamiento es DBSCAN que tiene un coeficiente de silhouette de 67.92% en comparación a las demás técnicas empleadas.
- Se ha validado el modelo considerando el informe climatológico de INNOCAR del año 2021 los patrones de comportamiento son similares a los encontrados en este estudio.

RECOMENDACIONES

- Se recomienda utilizar estas técnicas Machine Learning con datos observados recolectadas por el INHAMI para observar la efectividad y dar otro punto de vista diferente de patrones de comportamientos de precipitación en el Ecuador.
- Se recomienda experimentar con varias técnicas de aprendizaje automático para poder observar cuál de estas técnicas es adecuado para la base de datos en estudio.
- Se recomienda acudir donde un experto en el área de climatología para una mejor interpretación y comprensión de los eventos climatológicos que algunas veces no son tan evidentes en los análisis de datos.

BIBLIOGRAFÍA

AGUILERA, J. *Geografía General I. Geografía Física*. 1ª ed. España : UNED, 2020, pp. 34-42.

AGUILAR, S. *Evidencias del cambio climático en Ecuador*. 1ª ed. Ecuador, 2020, Agroecosistemas Revista para la transformación agraria sostenible, Vol. 8, pp. 72-76.

ALADINES, G. "Fenómeno del Niño historia y perspectivas". *Universidad de Cuenca: Facultad de Ciencias Médicas*, vol. 33, nº 3 (2015), (Ecuador) pp. 100-1009.

ANCAPICHÚN, S. "Variabilidad del Anticiclón Subtropical del Pacífico Sudeste y su impacto sobre la temperatura superficial del mar frente a la costa centro-norte de Chile". *Ciencias Marinas*, vol. 41, nº 1 (2015), (Chile) pp. 1-20.

ANKUSH, K. Y OTROS, 2020. Physics Guided Machine Learning Methods for Hydrology. *Física Atmosférica y Oceánica*, I(1), pp. 1-5.

ARMIJOS D, LOJÁN, 2010. Herpetofauna de un Bosque Húmedo Tropical en la Quinta "El Padmi" del Centro de Estudios y Desarrollo para la Amazonía (CEDAMAZ), provincia de Zamora Chinchipe. *Cedamaz*. 1(Primero): p. 92.

BOTERO, E. "El cambio climático y sus efectos en la biodiversidad en América Latina". *CEPAL*, vol. 2, nº 1 (2015), pp. 1-84.

BOBADILLA, J. *Machine Learning y Deep Learning: Usando Python, Scikit y Keras*. Colombia: Ediciones de la U, 2021, pp. 32-65.

BRAVO, A. "Reducción de Dimensiones: Revisión y Aplicaciones en Clasificación Automática". [En línea], 2020. [Consulta: 02 de septiembre 2022]. Disponible en: <http://uvadoc.uva.es/handle/10324/44147>.

BRENES, A y SABORÍO, V. "Elementos de climatología: su aplicación didáctica a Costa Rica". nº 1 (2015). Costa Rica : EUNED, 1995, pp. 37-112.

BUTLER, M. "Cartografía de recursos marinos: un manual de introducción". Roma-Italy : Food & Agriculture Org., 1990, pp. 12-34.

CASTILLO, M y JORDÁN, M. *Meteorología y clima*. Univ. Politèc. de Catalunya. n° 1 (1999). España, pp. 54-100.

CORTEZ, J. "Clustering de recetas culinarias generadas por giuseppe" vol. 1, n° 1 (2020), pp. 56-60.

CONTRERAS , A. "Espacio geográfico". *UAEN*, vol. 2, n° 3 (2001). México, pp. 13-32.

CONCEPTO, 2021. Editorial Etecé. [En línea] 2021. [Citado el: 13 de 01 de 2022.] <https://concepto.de/densidad/>.

CUADRAS, CARLES. *Nuevos métodos de análisis multivariante.*. España-Barcelona : CMC Editions, 2005, pp. 308-310.

DE LA FUENTE, S. *Análisis de Conglomerados*. España, 2011. pp. 57-100.

DOBILAS, S. "UMAP Dimensionality Reduction — An Incredibly Robust Machine Learning Algorithm". [En línea], 2021. [Consulta: 09 de febrero 2022]. Disponible en: <https://towardsdatascience.com/umap-dimensionality-reduction-an-incredibly-robust-machine-learning-algorithm-b5acb01de568>.

EL COMERCIO. EL COMERCIO. [En línea], 2020. [Consulta el: 28 de noviembre 2021]. Disponible en: <https://www.pressreader.com/ecuador/el-comercio-ecuador/20200126/page/1>

FERNÁNDEZ , J. "El cambio climático: Sus causas y efectos medioambientales". *Anuales de la Real Academia de Medicina y Cirugía de Valladolid Primero*, vol. 50, 2013, pp. 72-74.

HEREDIA, L. y ESCOBAR, Y. *Conglomerados*, España: Redalyc, vol. 1, 2012. pp. 11-50.

HERAS, J. *Clustering (Agrupamiento)* [blog], 2020. [Consulta: 19 de noviembre 2021]. Disponible en: <https://www.iartificial.net/clustering-agrupamiento-kmeans-ejemplos-en-python/>.

INAMHI. *El Instituto Nacional de Meteorología e Hidrología (INAMHI)* [blog], 2020. [Consulta: 16 de enero 2022]. Disponible en: <http://www.inamhi.gob.ec/boletin-especial-nevadas-21-de-agosto/>.

JUPYTER. *Project Jupyter* [blog], 2022. [Consulta: 17 de enero 2022]. Disponible en: <https://jupyter.org>.

KODINARIYA, T. y MAKWANA, P. *Review on Determining of Cluster in K-means Clustering.* International Journal of Advance Research in Computer Science and Management Studies, vol. 1, 2013, pp. 90-95.

LARA, U. Universidad Guayaquil. [En línea], 2022. [Consulta: 27 de Octubre 2021.]. Disponible en: <http://repositorio.ug.edu.ec/handle/redug/63075>.

LAUER, B. *Die niederschlagsjahreszeiten in ecuador und ihre klimadynamische interpretation.* Segunda. Washington: Erdkunde, 1992.

LOAIZA, A. Infobae. [En línea] 2021. [Consulta: 10 de Noviembre 2022.]. Disponible en: <https://www.primicias.ec/noticias/tecnologia/caf-ciudades-ecuador-enfrentar-cambio-climatico/>.

LÓPEZ, M. *Energía eólica.* España-Barcelona : Grupo Planeta (GBS), 2007, pp. 67-74.

MARTÍNEZ, E. E Libro. [En línea] 2020. [Consulta: 05 de enero 2022]. Disponible en: <https://elibro.net/es/ereader/esepoch/175596>.

MANDERY, C y GERLICH, N. *Welcome to Python* [blog]. 2021, [Consulta: 17 de enero 2022]. Disponible en: <https://www.python.org/success-stories/python-for-financial-machine-learning-at-union-investment/>.

MATUS, R. Estadística. [En línea], 2010. [Consulta: 05 de enero 2022] <https://elibro.net/es/ereader/esepoch/76119>.

MONIN, A. *An Introduction to the Theory of Climate.* Tokio, 1986, pp. 34-67.

MUÑOZ, D. "Manual de estadística". Bosnia : Juan Carlos Martínez Coll, 2000.

NADAL, I. y MUÑUZURI, V. *Mu/6-fundamentos de meteorología.* España : Univ Santiago de Compostela, 2006, pp. 23-29.

OCLES. France 24. [En línea], 2022. [Consulta: 10 de Noviembre 2022.]. Disponible en :<https://www.france24.com/es/minuto-a-minuto/20220409-seis-meses-de-lluvias-dejan-57-muertos-y-110-heridos-en-ecuador>.

OG. Observatio de la Globalización . [En línea] 2007. [Consulta: 22 de Diciembre de 2021.]. Disponible en :
https://flacsoandes.edu.ec/web/imagesFTP/1269363645.Honty_America_Latina_frente_al_cambio_climatico.pdf.

OMM. *Índice normalizado de precipitación*. Suiza : OMM, 2012.

OPS. Organización Panamericana de la Salud. [En línea] 2019. [Consulta: 10 de Noviembre 2022.]. Disponible en: <https://www.paho.org/es/temas/cambio-climatico-salud>.

PEÑA, D. *Análisis de datos multivariantes*. Madrid : McGraw-Hill Interamericana de España S.L., 2002. pp. 515-516.

PÉREZ, R. *Dinámica Atmosférica y los Procesos Tormentosos Severos*. España : Editorial Académica Española, 2006. pp. 365-368.

PETTERSEN, SVERRE. 1968. *Introducción a la meteorología* Baja California : UABC, 1968. pp. 69-70.

PICAZO, M. "Dinámica atmosférica y el clima". *Socllalluna* [En línea], 2021. [Consulta: 15 de enero 2022]. Disponible en: <https://socllalluna.com/2obach-ciencias-de-la-tierra-y-el-medio-ambiente/bloque-ii-las-capas-fluidas-dinamica/02-atmosfera-e-hidrosfera-dinamicas/04-capas-fluidas-y-el-clima/dinamica-atmosferica-y-el-clima/>.

PORTILLA, F. *Agroclimatología*. Quito : Universitaria Abya-Yala, 2018, pp. 68-72.

PORTILLO, G. "Meteorología en Red". *Meteorologiaenred* [En línea], 2017. [Consulta: 15 de enero 2022]. Disponible en: <https://www.meteorologiaenred.com/las-precipitaciones.html>.

RON, S. *BIOWEB* [blog], España. 2018. [Consulra: 31 de enero 2019]. Disponible en: <https://bioweb.bio/faunaweb/amphibiaweb/GeografiaClima/>.

SARKAR, D. y SHARMA, T. *Practical Machine Learning with Python*. . India : Apress, 2017. pp. 56-58.

SMITH, N. y PIERCE, C. *Resolución de problemas de química general*. Reverte, 1991. pp. 132-133.

TEJADA, A. *La humedad en la atmósfera bases físicas, instrumentos y aplicaciones*. México : Ucol, 2018. pp. 81-82.

TELÉGRAFO. El Telégrafo [en línea], 2019. [Consulta: 16 de enero 2022]. Disponible en: <https://www.eltelegrafo.com.ec/noticias/sociedad/6/ecuador-aumento-temperatura>.

TOULKERIDIS, T. "Climate Change according to Ecuadorian academics–Perceptions versus facts". vol. 31, (2020), pp. 21-46.

TURING, V. *Python* [blog]. [Consulta: 13 de febrero 2022]. Disponible en: <https://fragmentospython.blogspot.com/2020/05/ejemplo-calculo-de-distancias.html>.

VÁSQUEZ, A. Bits & Bricks. [blog]. [Consulta: 09 de febrero 2022]. Disponible en: <https://bitsandbricks.github.io/post/dbscan-machine-learning-para-detectar-centros-de-actividad-urbana/>.

VEGARA, J. *El cambio climático*. Barcelona : "la Caixa", 2010. pp. 76-81.

VIOLANTE, A. "An Introduction to t-SNE with Python Example". [En línea], 2018. [Consulta: 07 de febrero 2022]. Disponible en: <https://towardsdatascience.com/an-introduction-to-t-sne-with-python-example-5a3a293108d1>.

VIRGINIA. Desastres y sociedad. *El niño en america latina*. 13 de Diciembre de 1998, págs. 1-74.

YAN, Y y YAN, J. *Hands-On Data Science with Anaconda: Utilize the right mix of tools to create high-performance data science applications*. Mumbai : Packt Publishing Ltd, 2018. pp. 54-57.

ZÚÑIGA, A y CRESPO, A. *Meteorología y Climatología*. Primera. España : UNED, 2021.

ZHANG, R. RUOCHI.AI. [En línea], 2020. [Consulta: 09 de febrero 2022]. Disponible en:
<https://zhangruochi.com/DBSCAN-Clustering/2020/04/14/index.html>.



ANEXO

ANEXO A: AVAL DE GRUPO DE INVESTIGACIÓN GEAA



Riobamba, 02 de septiembre 2021

Señor

Miranda Sayay Wilson Geovanny

De mi consideración

Por la presente notifico a usted, que el Grupo de investigación GEAA 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 PRECIPITACIÓN 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
Fecha: 2021.09.02 20:08:53
-05'00'
Dr. Celso Guillermo Recalde
DIRECTOR GEAA-FAC. CIENCIAS

Teléfono: 593(03) 2998-200 Ext.2221

geaaepoch@gmail.com



epoch

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: 05 / 01 / 2023

INFORMACIÓN DEL AUTOR/A (S)
Nombres – Apellidos: Wilson Geovanny Miranda Sayay
INFORMACIÓN INSTITUCIONAL
Facultad: Ciencias
Carrera: Carrera Estadística
Título a optar: Ingeniero en Estadística Informática
f. responsable: Ing. Cristhian Fernando Castillo Ruiz


D.B.R.A.
Ing. Cristhian Castillo



2390-DBRA-UTP-2022