







Aprendizaje por refuerzo como soporte a la predicción de la precipitación mensual. Caso de estudio: Departamento de Boyacá - Colombia

Reinforcement Learning to Support Monthly Precipitation Prediction. Case Study: Department of Boyacá - Colombia

 Jimmy Alejandro Zea Gutiérrez¹;
  Marco Javier Suárez Barón²;
 Juan Sebastián González Sanabria³

¹ Universidad Pedagógica y Tecnológica de Colombia,
Sogamoso-Colombia,
jimmy.zea@uptc.edu.co

² Universidad Pedagógica y Tecnológica de Colombia,
Sogamoso-Colombia,
marco.suarez@uptc.edu.co

³ Universidad Pedagógica y Tecnológica de Colombia,
Tunja-Colombia,
juansebastian.gonzalez@uptc.edu.co

Cómo citar / How to cite

J. A. Zea Gutiérrez, M. J. Suárez Barón, J. S. González Sanabria,
“Aprendizaje por refuerzo como soporte a la predicción de la precipitación
mensual. Caso de estudio: Departamento de Boyacá-Colombia,”
Tecnológicas, vol. 27, no. 60, e3017, Jun. 2024.
<https://doi.org/10.22430/22565337.3017>

Resumen

La emisión de gases de efecto invernadero, atribuida directa o indirectamente a la actividad humana, es la principal causa del cambio climático a nivel mundial. Entre los gases emitidos, el dióxido de carbono (CO²) es el que más contribuye a la variación espacio temporal de magnitudes físicas como la humedad relativa, la presión atmosférica, la temperatura ambiente y, de manera más significativa, la precipitación. El objetivo de la investigación fue presentar un análisis de la predicción de la precipitación mensual en el departamento de Boyacá mediante el uso de modelos basados en aprendizaje reforzado (RL, por sus siglas en inglés). La metodología empleada consistió en extraer datos desde CHIRPS 2,0 (Climate Hazards Group InfraRed Precipitation with Station data, versión 2,0) con una resolución espacial de 0,05° que posteriormente fueron preprocesados para la implementación de enfoques basados en una simulación Montecarlo y aprendizaje reforzado profundo (DRL, por sus siglas en inglés) para proporcionar predicciones de la precipitación mensual. Los resultados obtenidos demostraron que la simulación Montecarlo como el DRL generan predicciones significativas de la precipitación mensual. Es esencial reconocer que los modelos convencionales basados en Aprendizaje profundo, como Memoria a Corto Plazo (LSTM) o Redes Convolucionales a Corto Plazo (ConvLSTM), pueden superar a los enfoques de simulación Montecarlo y DRL en términos de precisión de predicción. Se concluye que la implementación de técnicas de aprendizaje por refuerzo en modelos de predicción de la precipitación mensual detecta patrones de información que pueden ser usados como soporte a estrategias dirigidas a mitigar los riesgos económicos y sociales derivados de fenómenos climáticos.

Palabras clave

Aprendizaje automático, aprendizaje reforzado, modulación CHIRPS, simulación Montecarlo.

Abstract

The emission of greenhouse gases, directly or indirectly attributed to human activity, is the main cause of global climate change. Among the gases emitted, carbon dioxide (CO²) is the most important contributor to the spatio-temporal variation of physical quantities such as relative humidity, atmospheric pressure, ambient temperature and, most significantly, precipitation. The objective of the research was to present an analysis of the prediction of monthly precipitation in the department of Boyacá using models based on reinforced learning (RL). The methodology used consisted of extracting data from CHIRPS 2.0 (Climate Hazards Group InfraRed Precipitation with Station data, version 2.0) with a spatial resolution of 0.05 ° that were subsequently preprocessed for the implementation of Monte Carlo simulation and deep reinforced learning (DRL) approaches to provide monthly precipitation predictions. The results obtained showed that Monte Carlo simulation such as DRL generate meaningful predictions of monthly precipitation. It is essential to recognize that conventional models based on Deep Learning, such as Short-Term Memory (LSTM) or Short-Term Convolutional Networks (ConvLSTM), can outperform Monte Carlo and DRL simulation approaches in terms of prediction accuracy. It is concluded that the implementation of reinforcement learning techniques in monthly precipitation prediction models detects information patterns that can be used to support strategies aimed at mitigating economic and social risks derived from climate phenomena.

Keywords

Machine Learning, reinforcement learning, CHIRPS modulation, Monte Carlo simulation.

1. INTRODUCCIÓN

El análisis y pronóstico de patrones climáticos es un factor esencial en la comprensión y anticipación de las variaciones en las condiciones atmosféricas, lo que tiene repercusiones significativas en diversos sectores como: La agricultura, la gestión del agua y la planificación urbana [1]. Comprender y analizar los patrones de información en la precipitación a lo largo del tiempo permite dar respuesta a interrogantes como: ¿Cuáles son los meses más húmedos y secos en la región? ¿Cómo se distribuye la precipitación en diferentes áreas geográficas? ¿Se pueden identificar tendencias climáticas en la región? [2].

A pesar de la importancia de la predicción de la precipitación, en el departamento de Boyacá existen brechas de conocimiento relacionadas a la aplicación de modelos de pronóstico apoyados en técnicas computacionales. Los enfoques tradicionales, como los modelos de predicción actuales [3], a menudo adolecen de limitaciones para capturar patrones climáticos complejos y variaciones temporales en los datos. Además, la falta de herramientas interactivas de visualización puede dificultar la comunicación efectiva de los resultados a las partes interesadas.

En el ámbito de la predicción de variables cuantitativas, la aplicación del aprendizaje por refuerzo ha sido tradicionalmente una vía menos transitada [4]. A pesar de los avances significativos del RL para problemas como el control de juegos y la toma de decisiones en entornos complejos, su implementación en la predicción de variables numéricas ha sido un terreno en gran parte inexplorado [5].

Sin embargo, la creciente necesidad de pronósticos precisos en una variedad de campos [6], desde la meteorología hasta las finanzas, ha planteado la cuestión de si el RL podría ser una herramienta efectiva para mejorar la calidad de estas predicciones. Esta investigación se adentra en este territorio relativamente inexplorado y busca evaluar el potencial del RL en la predicción de variables cuantitativas, centrándose en un contexto particular: la predicción de la cantidad de precipitación mensual en el departamento de Boyacá. Esta sección, analiza la literatura más relevante que aportó para el desarrollo de este proyecto.

En [7] se presenta un enfoque innovador para mejorar la precisión en la predicción de lluvias costeras, subrayando su importancia en el contexto del cambio climático y sus impactos en la vida cotidiana y la economía. Se resalta la eficacia de las redes LSTM y el aprendizaje reforzado profundo en la predicción de lluvias, además de proponer un enfoque basado en Internet de las cosas (IoT, por sus siglas en inglés) que alcanza una precisión del 89 % en la predicción de lluvias costeras. El enfoque está centrado en regiones costeras, lo que podría limitar su aplicabilidad en áreas interiores o con diferentes características climatológicas. La precisión del 89 % se basa en el uso de IoT, lo que puede no ser replicable en regiones con infraestructura tecnológica limitada.

En [8] se introduce un modelo de pronóstico basado en series temporales que aborda la imputación de datos faltantes y la selección de variables clave. El modelo propuesto, basado en Bosques Randómicos del inglés Random Forest, supera a los modelos convencionales en términos de precisión de pronóstico. Con este artículo se pudo comprender de mejor manera el comportamiento de las series temporales para poder crear un ambiente para el modelo de DRL.

El modelo anterior está diseñado específicamente para el embalse Shimen, lo que puede limitar su aplicabilidad a otros cuerpos de agua con características diferentes. La necesidad de imputación de datos faltantes puede introducir incertidumbre adicional y afectar la precisión del modelo. Aunque se hace una selección de variables clave, existe el riesgo de que variables importantes no identificadas puedan influir en la precisión del pronóstico.

De otro lado, [9] se centra en la predicción de la velocidad del viento a corto plazo en sistemas de energía eólica. Propone un enfoque que combina la descomposición de datos, el aprendizaje por refuerzo profundo y la corrección de errores para mejorar la precisión de las predicciones. El estudio demuestra que este modelo puede capturar con precisión las fluctuaciones en los datos de velocidad del viento, con aplicaciones potenciales en la industria de la energía eólica. El modelo en mención se centra en la predicción a corto plazo, lo que podría no ser adecuado para aplicaciones que requieren pronósticos a largo plazo. La precisión del modelo depende de la efectividad de la descomposición de datos y la corrección de errores, lo que puede variar según las características de los datos de entrada. El enfoque está diseñado específicamente para la velocidad del viento, lo que puede limitar su transferencia directa a otros fenómenos meteorológicos.

La aplicación del aprendizaje profundo basado en refuerzo en estrategias de control en tiempo real para sistemas de drenaje urbano [10]. Se destaca en este trabajo que, a pesar de las incertidumbres en las señales de medición, las estrategias de DRL son robustas y efectivas en la mejora de la eficiencia de los sistemas de drenaje urbano, especialmente en el contexto de eventos climáticos extremos más frecuentes debido al cambio climático [11]. Se resalta la superioridad de las estrategias de DRL en comparación con las basadas en reglas. A pesar de la robustez del modelo, la presencia de incertidumbres en las señales de medición puede afectar su rendimiento. La eficacia del modelo se prueba principalmente en eventos climáticos extremos, lo que podría limitar su rendimiento en condiciones meteorológicas más moderadas. El aprendizaje automático ha demostrado ser eficaz en la predicción de patrones climáticos en otras regiones [12] y se busca aplicarlo a datos de precipitación mensual en Boyacá, evaluando su capacidad predictiva. El presente trabajo analiza los desafíos de la problemática del cambio climático mediante el desarrollo y evaluación de modelos de predicción basados en el aprendizaje reforzado.

Este proyecto integra dos contextos: En primer lugar, el desarrollo de modelos de pronóstico a la precipitación mensual en el departamento de Boyacá utilizando simulación Montecarlo y aprendizaje reforzado profundo; y en segundo lugar la creación de una herramienta interactiva de visualización que tiene el propósito de facilitar la exploración y la comunicación de los resultados del pronóstico a usuarios interesados [13], facilitando la toma de decisiones y una comprensión de los patrones climáticos obtenidos.

En cuanto a la adquisición de datos, se extrajeron registros históricos de precipitación del conjunto de datos CHIRPS 2,0 con una resolución espacial de 0,05°. La elección de esta fuente de información se basó en la calidad y disponibilidad de los datos, factores fundamentales para asegurar la mayor exactitud de las predicciones. Las técnicas y algoritmos empleados, se optó por implementar Simulación Montecarlo y Aprendizaje Reforzado Profundo (DRL). Esta decisión se tomó con el propósito de evaluar la capacidad predictiva de algoritmos basados en aprendizaje por refuerzo [14] para su posterior comparación con algoritmos de aprendizaje profundo. Los métodos anteriores permitieron capturar patrones en la lluvia, lo cual fue esencial para generar pronósticos certeros.

La contribución principal de este artículo se enfoca en la aplicación de modelos basados en aprendizaje reforzado para predecir la lluvia mensual en el departamento de Boyacá. Para lograrlo, se evaluó el rendimiento de estos modelos en términos de métricas de precisión y capacidad predictiva [15], y se proporcionó una representación visual interactiva de los datos históricos y las predicciones a través de un tablero de visualización. Finalmente, la selección de métricas de evaluación apropiadas, como MAE, MSE y RMSE, desempeñó un papel esencial en el proyecto. Estas métricas se eligieron con el fin de medir objetivamente la precisión y capacidad predictiva de los modelos [16], proporcionando una base sólida para evaluar su rendimiento.

2. ASPECTOS METODOLÓGICOS

2.1 Caso de estudio

El presente proyecto se desarrolla en el departamento de Boyacá, ubicado en la región Andina del centro de Colombia. Boyacá es reconocido por su actividad agrícola, lo que subraya la relevancia de pronosticar la precipitación en la toma de decisiones relacionadas con la agricultura y la gestión de recursos hídricos en la región [17]. La Figura 1 muestra con claridad la ubicación geográfica de Boyacá en el contexto de Colombia.



Figura 1. Ubicación geográfica del departamento de Boyacá. Fuente: elaboración propia.

2.2 Flujo de información

El proceso se inicia con la extracción de archivos climáticos de la base de datos CHIRPS 2,0, abarcando registros de precipitación desde 1981 hasta mayo de 2023, con una resolución espacial de 0,05°. La preparación de los datos incluye la consolidación de registros a nivel mensual y validación de valores atípicos [18].

A continuación, se procede a la integración de las bases de datos y se lleva a cabo un filtro geográfico para obtener los registros específicos de la región de Boyacá. En el proceso de selección de características, se consideran la latitud, longitud y la fecha (año y mes) como elementos fundamentales para capturar las variaciones locales y temporales relevantes en el pronóstico de precipitación [19].

Por último, se aplican modelos de aprendizaje reforzado basados en simulación Montecarlo y aprendizaje reforzado profundo para predecir la precipitación mensual. Estos modelos se entrenan y validan utilizando métricas como el Error Absoluto Medio (MAE, por sus siglas en inglés) y el Error Cuadrático Medio (MSE, por sus siglas en inglés).

2.2.1 Adquisición de datos

En esta tarea se realiza la descarga de archivos desde el repositorio web que contiene enlaces a archivos climáticos en formato netCDF (.nc). Se utilizan bibliotecas como requests y BeautifulSoup para solicitar la página web y analizar su contenido. Luego, se descarga cada archivo identificado en los enlaces y se almacenan.

Este trabajo utiliza datos de precipitación de CHIRPS 2,0 [20], una fuente reconocida en investigación climática a nivel global. Los datos, con una resolución espacial de 0,05°, abarcan desde 1981 hasta mayo de 2023. Se preprocesaron consolidando los valores de precipitación mensualmente mediante la suma de datos diarios, lo que permite un análisis más relevante de tendencias climáticas a lo largo de los meses. Para ello, se descargaron datos diarios desde 1981 y se combinaron en archivos netCDF (.nc) [21] para facilitar el procesamiento.

2.2.2 Conformación del dataset

Como se ha mencionado, se hizo uso de CHIRPS 2,0, estos datos incluyen información importante como la ubicación geográfica, cantidad de precipitación y la fecha de registro [22]. Para adaptar estos datos a la región de Boyacá, se aplicaron técnicas de filtrado geográfico y se utilizaron coordenadas geográficas para delimitar la zona de interés. Además, se empleó la API de Google Maps para relacionar los datos con los municipios de Boyacá, permitiendo crear un conjunto de datos más específico analizar la precipitación mensual en las cuencas hidrográficas de este departamento. En la Figura 2 se puede observar lo anteriormente mencionado.

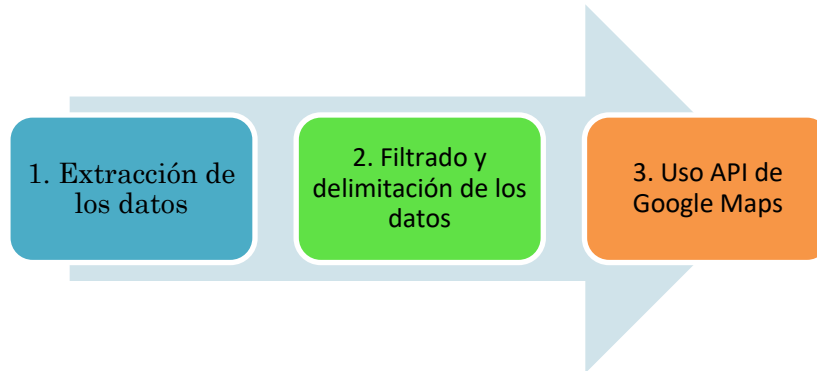


Figura 2. Conformación del conjunto de datos. Fuente: elaboración propia.

2.2.3 Preprocesamiento

En la Figura 3 se observa el flujo del preprocesamiento de datos el cual fue una fase crucial en la preparación de los datos climáticos para el análisis subsiguiente. Aunque no se encontraron valores faltantes en el conjunto de datos, se identificaron valores "atípicos" en la variable de precipitación. Sin embargo, se decidió no intervenir en estos valores atípicos debido a su importancia en el estudio.

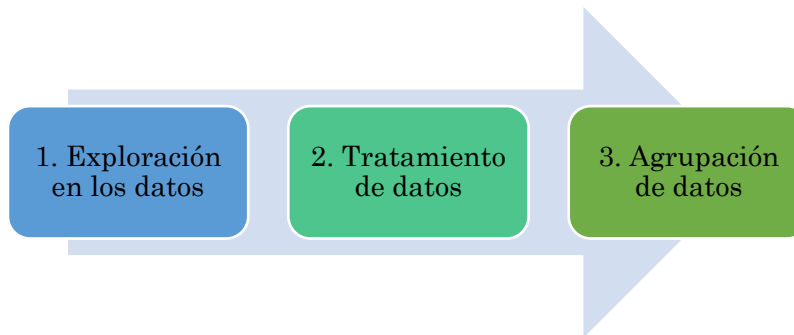


Figura 3. Preprocesamiento de datos. Fuente: elaboración propia.

Para detectar los valores atípicos, se utilizaron diagramas de boxplot aplicados a cada mes y a cada uno de los puntos de medición de precipitación. Estos diagramas identifican valores atípicos utilizando los cuartiles y el rango intercuartil (IQR) de la siguiente manera:

- Cálculo del IQR: El IQR se calcula como la diferencia entre el tercer cuartil (Q3) y el primer cuartil (Q1).
- Identificación de valores atípicos: Los valores se consideran atípicos si están por debajo de $Q1 - 1.5 * IQR$ o por encima de $Q3 + 1.5 * IQR$.

En la Figura 4 se muestra un ejemplo de los valores atípicos encontrados para los puntos de medición del municipio de puerto Boyacá en 2022.

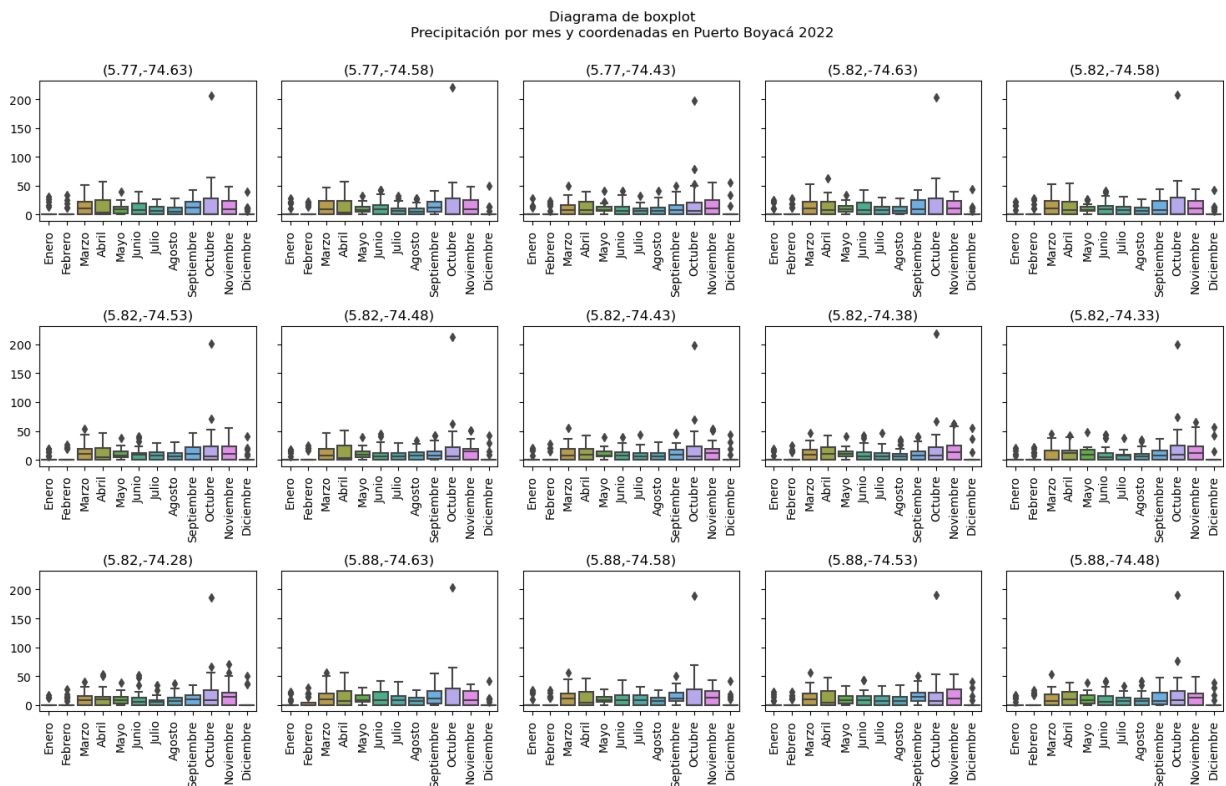


Figura 4. Boxplot valores atípicos puerto Boyacá 2022. Fuente: elaboración propia.

El rango de estos valores atípicos variaba dependiendo del mes y del punto de medición de la precipitación, pero por lo general el mes que presentaba el rango más alto en los valores atípicos era octubre, en el ejemplo de la Figura 4 se observa que en dicho mes la media de precipitación no superaba los 50 mm diarios mientras que los valores atípicos alcanzaban casi los 200 mm diarios.

La variable de precipitación es propensa a presentar valores extremos durante eventos climáticos intensos, que son precisamente el foco de estudio. Estas "anomalías" son valiosas fuentes de información y se consideraron legítimas y relevantes para el análisis, ya que reflejan de manera realista condiciones climáticas excepcionales en la región de Boyacá. Cualquier manipulación de estos valores podría distorsionar los datos y afectar negativamente los resultados y conclusiones.

Antes de aplicar los algoritmos, los datos se agruparon por latitud, longitud, municipio, año y mes, y se sumaron los valores de precipitación, ya que las predicciones se realizan para

precipitación total mensual en coordenadas y meses específicos. Las demás columnas se mantuvieron para proporcionar información adicional en los resultados.

La agrupación por municipio se basa en etiquetas de municipios asignadas a cada conjunto de coordenadas de medición de precipitación, y luego se suman los valores de precipitación mensual correspondientes a esas coordenadas y periodos de tiempo específicos por lo que se usa como etiqueta para el píxel.

2.2.4 Conjunto de datos obtenido

Una vez terminado el preprocesamiento, el conjunto de datos que se obtuvo fue alojado en [6], y consta de 11664723 registros y 7 columnas que representan variables previamente descritas. Estos datos abarcan la precipitación diaria desde enero de 1981 hasta mayo de 2023, con una resolución espacial de 0,05°. Se observan 753 puntos de medición distribuidos en todos los municipios de Boyacá, según se ilustra en las Figuras 5a y 5b, generadas a partir de información geoespacial actualizada hasta 2022.

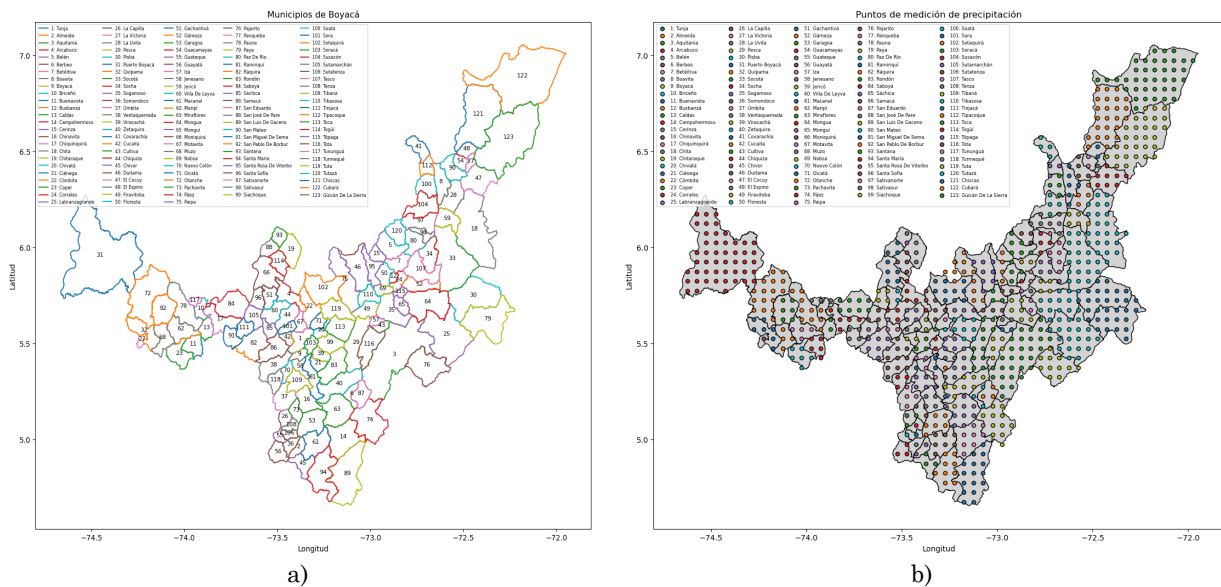


Figura 5. a) Municipios del departamento de Boyacá. b) Puntos de medición de precipitación. Fuente: elaboración propia.

2.3 Construcción de los modelos

En esta sección, se describe la metodología utilizada para la construcción de los modelos para el análisis climático, que emplea simulación Montecarlo y aprendizaje reforzado profundo [23]. Adaptando estas arquitecturas para predecir la precipitación en el departamento de Boyacá, siendo ampliamente utilizadas en el procesamiento secuencial y espacial de datos.

Para la construcción y evaluación de modelos predictivos, se dividió el conjunto de datos en tres conjuntos distintos: entrenamiento, validación y prueba. En este proceso, se asignó el 80 % de los datos disponibles para el conjunto de entrenamiento, mientras que el 20 % restante se reservó para el conjunto de validación, para estos dos conjuntos no se incluyeron datos de los últimos tres años ya que estos datos fueron usados para el conjunto de datos de prueba.

El conjunto de entrenamiento para ajustar los diferentes parámetros y arquitecturas del modelo, mientras que el conjunto de validación se empleó para validar y refinar estos modelos. En el período de calibración, que se realizó con el conjunto de entrenamiento, se exploraron diversas configuraciones de modelos y se ajustaron sus parámetros utilizando técnicas como la validación cruzada y la búsqueda de hiperparámetros. Esta fase permitió optimizar el rendimiento de los modelos en base a los datos disponibles, asegurando así una adaptación óptima a las características específicas de la precipitación en el departamento de Boyacá.

Posteriormente, se utilizó el conjunto de validación para evaluar la capacidad predictiva de los modelos entrenados y realizar ajustes adicionales según fuera necesario. Durante esta etapa, se llevó a cabo una validación para verificar la capacidad de generalización de los modelos y su capacidad para realizar predicciones precisas en datos no vistos.

Finalmente, en el período de prueba, se evaluó la calidad predictiva de los modelos utilizando el conjunto de datos de prueba, que no fue empleado en la calibración ni en la validación. Esta fase proporcionó una evaluación independiente y objetiva de la capacidad de los modelos para generalizar y realizar predicciones en situaciones del mundo real.

2.3.1 Selección de características

En el análisis de datos y la predicción de información, la caracterización y descripción de las diversas características que conforman los datos desempeñan un rol esencial. Cada atributo no solo representa una particularidad inherente a la información, sino que también contribuye de manera significativa a las conclusiones y predicciones derivadas del análisis. En el contexto de la predicción de precipitación mensual, la identificación y caracterización precisas de estas propiedades resultan cruciales para comprender las causas subyacentes que influyen en los patrones de precipitación, lo que en última instancia mejora la precisión de los modelos predictivos.

A continuación, se presenta la Tabla 1, que proporciona una descripción detallada de las variables utilizadas en la construcción de los modelos de predicción de precipitación. Cada variable se detalla en términos de su tipo de dato, naturaleza y se ofrece una breve descripción de su importancia en el contexto climático.

Tabla 1. Variables usadas en la construcción de los modelos. Fuente: elaboración propia.

Variable	Tipo de dato	Descripción
Latitud	Numérico	Coordenada norte/sur
Longitud	Numérico	Coordenada este/oeste
Tiempo	Fecha	Representa la fecha del registro de medición
Precipitación	Numérico	Cantidad de precipitación en milímetros

Cada uno de los atributos identificados tiene una función única en la construcción y validación de los modelos predictivos. Su selección y caracterización permiten comprender las relaciones complejas entre diferentes factores espacio temporales y cómo influyen en los patrones de precipitación. Estos atributos se consideran en conjunto en la creación de los modelos que puedan capturar la variabilidad y las tendencias en los datos de precipitación, lo que finalmente conduce a predicciones más precisas y fundamentadas.

Las características del pronóstico en cada uno de los modelos empleados son mensuales. El "timestep" del pronóstico es de 1 mes, ya que las predicciones se realizan para la precipitación total mensual en coordenadas y meses específicos.

2.3.2 Simulación Montecarlo

En la implementación de la simulación Montecarlo para el proyecto, se adoptó un enfoque que se basa en la generación de un conjunto de resultados posibles utilizando distribuciones de probabilidad, en lugar de utilizar un conjunto fijo de valores de entrada. Este método recalcula repetidamente los resultados [24], utilizando diferentes números aleatorios dentro de un rango definido por valores mínimo y máximo. La simulación Montecarlo se destaca por su alta precisión, especialmente a medida que se aumenta el número de iteraciones, lo que permite proyectar con mayor exactitud resultados a largo plazo. Un ejemplo simple de su utilidad es el cálculo de probabilidades al lanzar dos dados, donde se obtienen predicciones más precisas al simular un mayor número de lanzamientos.

En este proyecto, la simulación Montecarlo se empleó para predecir la precipitación futura en diferentes coordenadas geográficas durante un mes específico, abarcando los meses 6 a 12, y permitiendo obtener valores de precipitación previstos para los últimos años de 2023. En la Figura 6 se presenta la arquitectura del modelo de simulación Montecarlo aplicado.

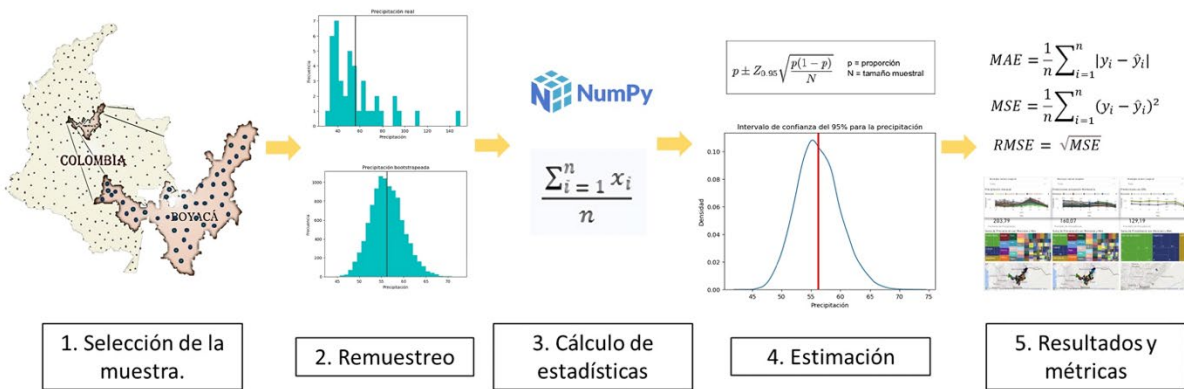


Figura 6. Flujo de datos de la simulación Montecarlo. Fuente: elaboración propia.

El proceso de aplicación de la simulación a los datos de precipitación consta de varios pasos claves descritos a continuación.

- Selección y filtrado de datos de interés: Se elige una coordenada geográfica y un mes específico para la predicción. Los datos originales se filtran para obtener fi-las que coincidan con estos criterios, creando un nuevo dataframe para la simulación.
- Remuestreo: Se emplea la simulación Montecarlo con el método de remuestreo bootstrap, realizándose 10000 iteraciones. En cada iteración, se realiza un remuestreo aleatorio con reemplazo de los valores reales de precipitación.
- Cálculo de estadísticas: Se extraen los valores reales de precipitación para la coordenada y mes seleccionados y se calcula el promedio de precipitación a partir de los datos re muestreados, generando una predicción posible de precipitación.
- Estimación: Luego de todas las simulaciones, se calcula un intervalo de confianza del 95 % para las predicciones, indicando la probable ubicación del valor real de precipitación con un nivel de confianza del 95 %. Un intervalo más estrecho implica mayor confianza en la predicción.
- Resultados y métricas de evaluación: Se evalúa el rendimiento del método Montecarlo calculando valores de precipitación promedio y creando un nuevo dataframe con

columnas que incluyen latitude, longitude, month, year (limitado a 2023), municipality y precip_predicted, que alberga las predicciones de precipitación.

Las distribuciones de probabilidad se ajustaron mediante la técnica de bootstrap. Esto implicó generar múltiples conjuntos de datos remuestreados a partir de los datos reales de precipitación y calcular la media de cada conjunto remuestreado. Así, se obtuvo una distribución de valores simulados de precipitación.

Por otro lado, el pronóstico de precipitación se tomó como el promedio de los valores simulados. Este valor promedio representa la estimación central de la precipitación mensual para las coordenadas y el mes específico. Y las distribuciones ajustadas son dependientes del mes a pronosticar. En cada iteración, se filtran los datos por mes y se ajustan las distribuciones de probabilidad específicamente para los datos de precipitación de ese mes. Esto permite capturar las características específicas de la precipitación en cada mes, reflejando la estacionalidad y variaciones mensuales.

2.3.3 Aprendizaje por refuerzo profundo

El segundo modelo de predicción de precipitación se basa en el aprendizaje reforzado profundo y utiliza un agente inteligente personalizado para tomar decisiones óptimas según los datos históricos. Se crea un entorno para entrenar al agente en la predicción de precipitación, donde el agente toma decisiones (0 para "sin lluvia" y 1 para "lluvia") basadas en una política que considera características relevantes como latitud, longitud, mes y año. El agente busca maximizar una recompensa acumulativa a lo largo del tiempo, aprendiendo de su experiencia sin recibir información explícita sobre cómo actuar en situaciones específicas.

En el contexto del aprendizaje por refuerzo profundo aplicado a la predicción de la precipitación, es importante destacar que el aprendizaje por refuerzo, en su esencia, no está centrado en la predicción de variables cuantitativas como lo es la cantidad de precipitación en milímetros. Por esta razón, se optó por utilizar el aprendizaje por refuerzo profundo, el cual emplea una red neuronal profunda para tomar decisiones óptimas basadas en datos históricos.

En este enfoque, cada acción se asigna a un valor simbólico: 0 para "sin lluvia" y 1 para "lluvia", reflejando así la naturaleza discreta de las decisiones en este contexto. El agente inteligente entrenado se enfrenta a un entorno único diseñado para la predicción de la precipitación, donde debe tomar decisiones sobre la probabilidad de lluvia en función de características relevantes como latitud, longitud, mes y año.

En este modelo se utilizan técnicas de optimización para descubrir la política óptima del agente, que relaciona los estados del entorno con sus acciones. Para ello, se emplea una red neuronal profunda para aproximar la función de valor de acción, lo que permite al agente manejar entornos complejos y conjuntos de datos extensos. La selección de la acción se basa en las expectativas de recompensa futura calculadas por la red neuronal, utilizando un algoritmo llamado Deep Q-Network (DQN) que se basa en técnicas avanzadas de aprendizaje profundo para estimar y actualizar en tiempo real la función de valor de acción.

A continuación, se detalla mejor tal proceso. Para alcanzar este objetivo, se emplea el aprendizaje reforzado con técnicas de optimización para descubrir la política óptima del agente, que relaciona los estados del entorno con sus acciones. El algoritmo de aprendizaje reforzado profundo utiliza una red neuronal profunda para aproximar la función Q, lo que le permite lidiar con entornos complejos y conjuntos de datos extensos. El agente selecciona la acción que se espera que tenga la recompensa más alta según las expectativas de recompensa futura calculadas por la red neuronal. El ciclo se repite para encontrar la política óptima que

maximice las recompensas acumuladas a largo plazo, utilizando un algoritmo llamado Deep Q-Network (DQN) que se basa en técnicas avanzadas de aprendizaje profundo para estimar y actualizar en tiempo real la función Q. En la Figura 7 se presenta la arquitectura del modelo de simulación Montecarlo aplicado.

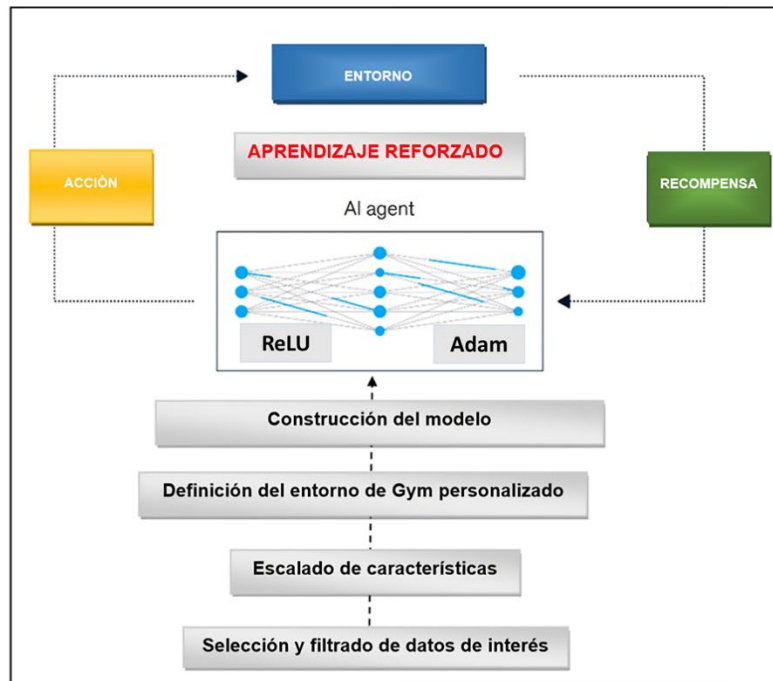


Figura 7. Flujo de datos DRL. Fuente: elaboración propia.

El proceso de aplicación del aprendizaje reforzado profundo para predecir la precipitación futura se describe a continuación [25].

- Selección y filtrado de datos de interés: Se realiza una selección y filtrado de datos similar al utilizado en la simulación Montecarlo.
- Escalado de características: Antes de aplicar el aprendizaje reforzado profundo, se escalan las características relevantes (latitud, longitud, mes y año) utilizando la función StandardScaler de Scikit-Learn. Esto normaliza las características y facilita el aprendizaje eficiente del modelo.
- Definición del entorno Gym personalizado: Se crea un entorno de Gym personalizado específicamente para el problema de predicción de precipitación. Este entorno incluye un espacio de observación para las características escaladas y un espacio de acción discreto (0 para "sin lluvia" y 1 para "lluvia").
- Construcción del modelo: Se construye un modelo de aprendizaje por refuerzo profundo utilizando TensorFlow y Keras. El modelo consta de una capa de entrada con 64 neuronas, una función de activación ReLU y una capa de salida única para predecir la cantidad de precipitación. El modelo se compila con el optimizador Adam y la función de pérdida de error cuadrático medio (MSE).

La elección de la función de activación ReLU (Rectified Linear Unit, por sus siglas en inglés) se basa en su capacidad para introducir no linealidad en el modelo al devolver el valor de entrada si es positivo y cero en caso contrario, lo que ayuda a capturar relaciones no lineales en los datos.

El optimizador Adam se selecciona debido a su eficacia en la mayoría de los problemas de optimización, ya que ajusta automáticamente la tasa de aprendizaje durante el entrenamiento para minimizar la función de pérdida y mejorar la precisión de las predicciones.

2.3.4 Entrenamiento del agente del modelo DRL

El entrenamiento del agente comprende 50 episodios, donde cada episodio representa una interacción entre el agente y el entorno (Figura 8). El proceso de entrenamiento consta de varias etapas.

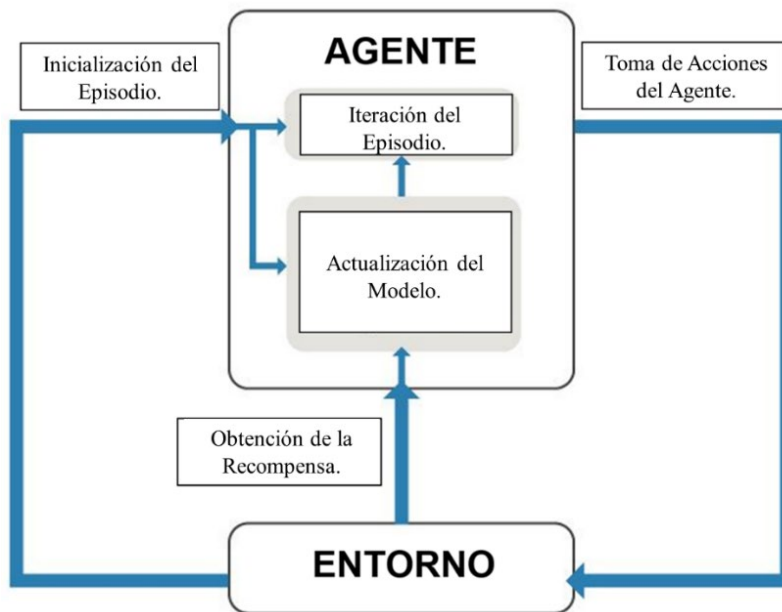


Figura 8. Entrenamiento del agente DRL. Fuente: elaboración propia.

- Inicialización del episodio: Al comienzo de cada episodio, el agente se reinicia, devolviendo el estado inicial del entorno con características escaladas (latitud, longitud, mes y año).
- Toma de decisiones del agente: El agente elige una acción (0: No llover, 1: Llover) según su política actual. Si la exploración está activada, se selecciona una acción aleatoria de lo contrario se elige la acción predicha por el modelo.
- Ejecución de la acción y recompensa: El agente ejecuta la acción en el entorno y obtiene el siguiente estado y la recompensa, que se calcula como la diferencia entre la predicción y el valor real de la precipitación, con un ajuste en caso de acción 0.
- Actualización del modelo: El modelo se actualiza utilizando la recompensa y el siguiente estado, con el objetivo de maximizar la recompensa acumulada y minimizar la diferencia entre predicciones y observaciones.
- Iteración del episodio: Estas etapas se repiten para cada paso de tiempo en el episodio. El agente toma decisiones secuenciales, ejecuta acciones y ajusta el modelo para aprender a hacer predicciones precisas de la precipitación que se asemejen a los datos reales.

2.3.5 Función de recompensa

La función de recompensa es fundamental en el aprendizaje por refuerzo y guía las decisiones del agente. En este proyecto, se define de la siguiente manera:

- Exploración vs explotación: Antes de calcular la recompensa, se considera un elemento clave en el aprendizaje por refuerzo: la exploración frente a la explotación. El agente tiene una probabilidad de tomar acciones aleatorias en lugar de seguir su política actual, fomentando la exploración y evitando patrones subóptimos.
- Cálculo de la recompensa: En cada paso de tiempo, la recompensa se calcula como la diferencia absoluta entre la predicción del modelo y la precipitación real observada. Si el agente elige la acción "sin lluvia" (acción 0), se le aplica una penalización adicional de -0,1, incentivando la elección de "lluvia" para maximizar la recompensa acumulada. La penalización se justifica porque no se registran valores de precipitación de 0 mm para ningún mes.
- Finalización de episodios: Cada episodio se da por completo cuando el agente ha explorado todos los pasos disponibles en el conjunto de datos, asegurando que haya aprendido de todas las observaciones disponibles.

2.4 Métricas de evaluación

Después de finalizar el preprocesamiento de datos y la creación de los modelos de pronóstico, es esencial llevar a cabo la validación y evaluación de estos modelos para garantizar que puedan ofrecer predicciones precisas de la precipitación mensual en el departamento de Boyacá. Este proceso incluye el uso de métricas destinadas a evaluar el desempeño y la eficacia de los modelos.

2.4.1 Error cuadrático medio (MSE)

Esta métrica convierte en una herramienta fundamental para evaluar la precisión de los modelos de predicción de precipitación. La fórmula (1) permite medir cuán cerca están las predicciones del modelo de las mediciones reales de precipitación en un conjunto de datos.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

Donde:

- MSE representa el Error Cuadrático Medio, que es un valor numérico que indica cuánto varían las predicciones del modelo con respecto a los datos reales de precipitación.
- n representa el total de observaciones en el conjunto de datos. Estas observaciones corresponden a mediciones de precipitación en diferentes ubicaciones y momentos.
- y_i representa el valor real de la precipitación en la observación i.
- \hat{y}_i es la predicción de la cantidad de precipitación que el modelo ha generado para la observación i.

El MSE proporciona una medida cuantitativa de cuán preciso es el modelo de predicción de precipitación. Cuanto menor sea el valor del MSE, más cercanas estarán las predicciones del modelo a los datos reales de precipitación. Esto significa que, si se minimiza el MSE, se

estaría mejorando la capacidad del modelo para predecir con precisión la cantidad de precipitación en diferentes ubicaciones y momentos.

2.4.2 Raíz del error cuadrático medio (RMSE)

En (2) se presenta la métrica que se deriva directamente del MSE y se utiliza para evaluar la calidad de las predicciones de los modelos. Esta métrica proporciona una medida más interpretable de la magnitud del error promedio entre las predicciones y los valores reales.

$$RMSE = \sqrt{MSE} \quad (2)$$

Donde:

- *RMSE* representa la raíz del MSE, que es un valor numérico que nos indica cuánto se desvían en promedio las predicciones del modelo de las mediciones reales de precipitación.
- MSE es el Error Cuadrático Medio, que se explicó previamente y que mide el promedio de los errores al cuadrado entre las predicciones y los valores reales.

2.4.3 Error absoluto medio (MAE)

El MAE presentado en (3) se ha calculado como una medida de la magnitud promedio de los errores entre las predicciones de nuestros modelos y los valores reales de precipitación mensual en Boyacá. Para realizar este cálculo, se ha utilizado la función `mean_absolute_error` de la biblioteca `sklearn.metrics`. El MAE se expresa de la siguiente manera:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

Donde:

- n es el número de observaciones.
- y_i representa el valor real de precipitación mensual.
- \hat{y}_i representa la predicción del modelo para el valor de precipitación mensual.

3. RESULTADOS

3.1 Análisis exploratorio

A partir del conjunto de datos, se calcula el porcentaje de estaciones de medición de precipitación, lo cual guarda una relación directa con el tamaño de cada municipio. En otras palabras, los municipios más extensos albergan una mayor cantidad de puntos de medición. De esta manera, se destaca que el municipio de Puerto Boyacá representa el 6,5 % del total de estaciones de medición, con 49 puntos de medición; seguido por Cubará y Aquitania en su orden, como se ilustra en la Figura 9.

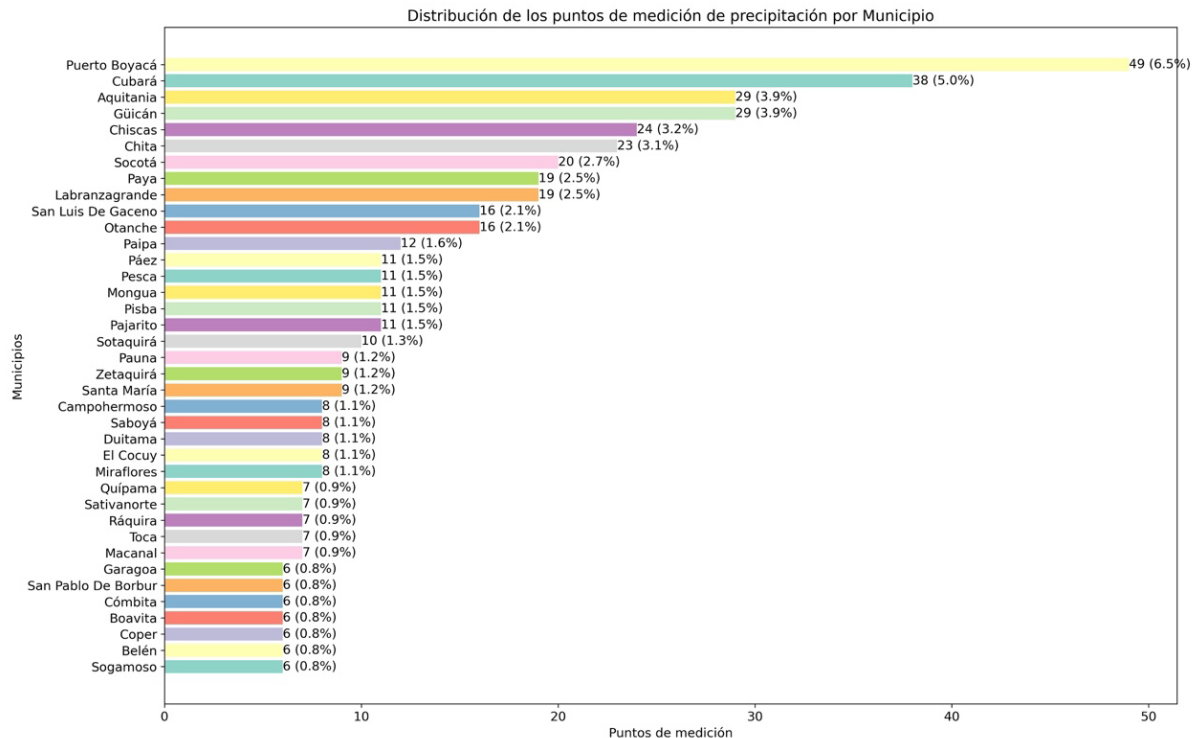


Figura 9. Distribución de coordenadas por municipios. Fuente: elaboración propia.

Los datos extraídos de CHIRPS 2,0, con una resolución espacial de 0,05 grados, revelan que la distancia entre estaciones de medición es de aproximadamente 5,53 kilómetros. Como ejemplo, la Figura 10 presenta la precipitación mensual de abril de 2022 en puntos de medición. Se emplea una escala de colores, donde tonos más oscuros de azul indican una mayor precipitación en un punto específico.

Por otro lado, la Figura 11 muestra la precipitación total mensual en cada punto dentro del municipio de Sogamoso durante el año 2022. Esta visualización proporciona información detallada sobre eventos específicos en un lugar particular, permitiendo conclusiones más precisas.

El análisis de la precipitación mensual en el departamento de Boyacá para el año 2022, representado en los mapas de calor de la Figura 12, ofrece información valiosa.

- La distribución espacial de la precipitación, destacando patrones de concentración de lluvia en diferentes áreas durante diversos meses. Se pueden identificar claramente los meses más húmedos, como mayo, agosto y octubre, con precipitaciones de hasta 400 mm, así como los meses más secos, como diciembre, enero y febrero.
- Se observa que la topografía del departamento de Boyacá influye en la distribución de la precipitación, ya que áreas montañosas, valles y llanuras exhiben patrones de precipitación distintos.
- La representación visual permite comparar la distribución de la precipitación entre los diferentes meses, lo que facilita la comprensión de cómo cambian las condiciones climáticas mes a mes.
- La visualización también busca identificar posibles anomalías climáticas, como meses inusualmente secos o húmedos en comparación con el patrón general. Sin embargo, en el caso del año 2022, según la Figura 12, no se observan anomalías significativas.

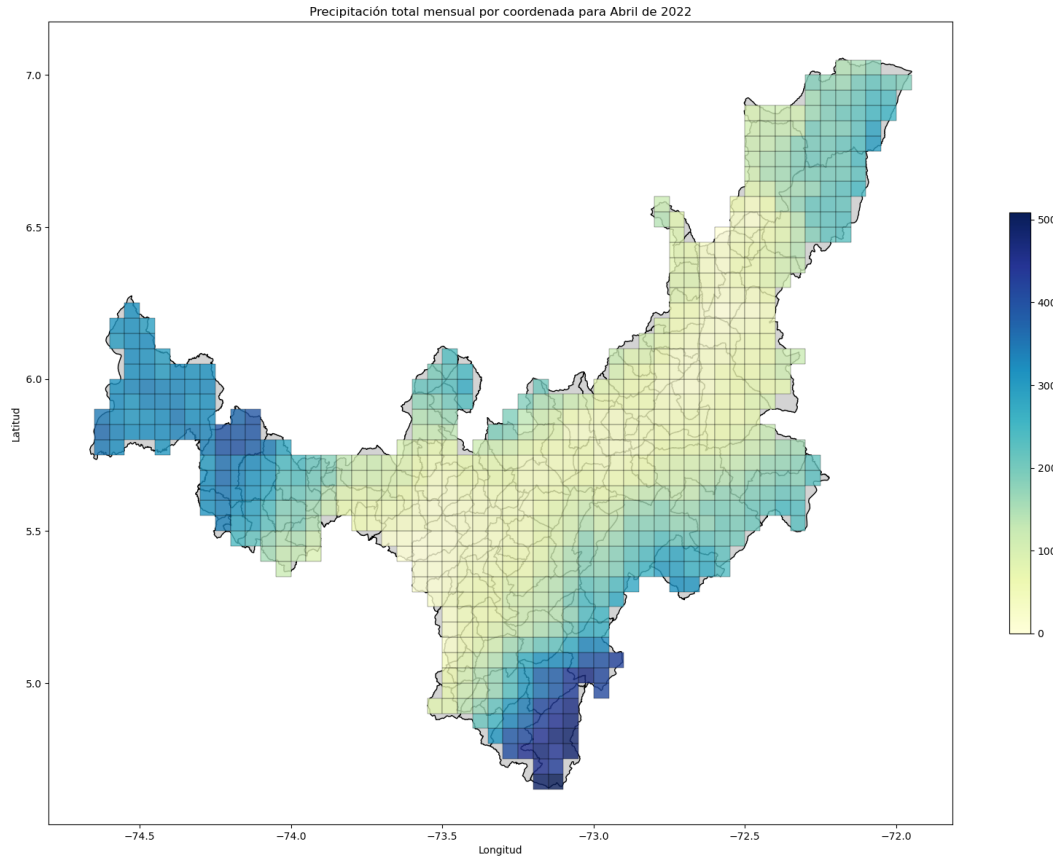


Figura 10. Precipitación mensual por puntos de medición abril 2022. Fuente: elaboración propia.

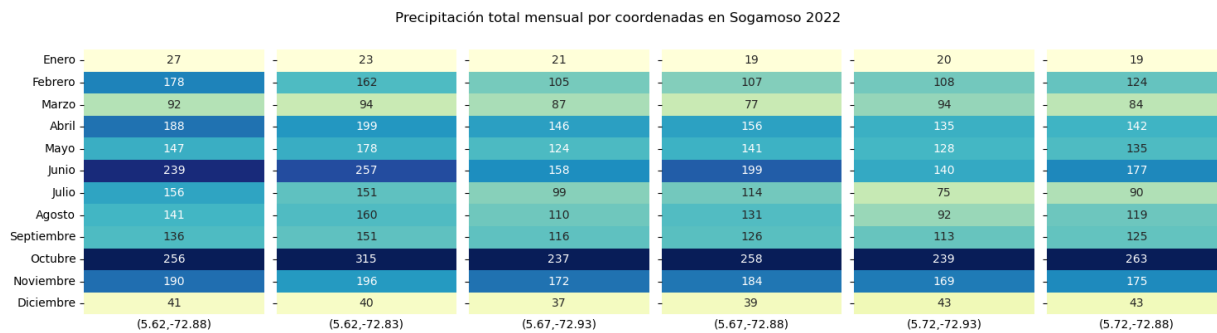


Figura 11. Precipitación total mensual por coordenadas en Sogamoso 2022. Fuente: elaboración propia.

En general, el análisis y visualización de la precipitación mensual por coordenadas permiten obtener información detallada sobre las condiciones climáticas locales y cómo varían a lo largo del tiempo, lo que puede ser útil para tomar decisiones informadas en áreas como la agricultura, la gestión del agua y la planificación urbana.

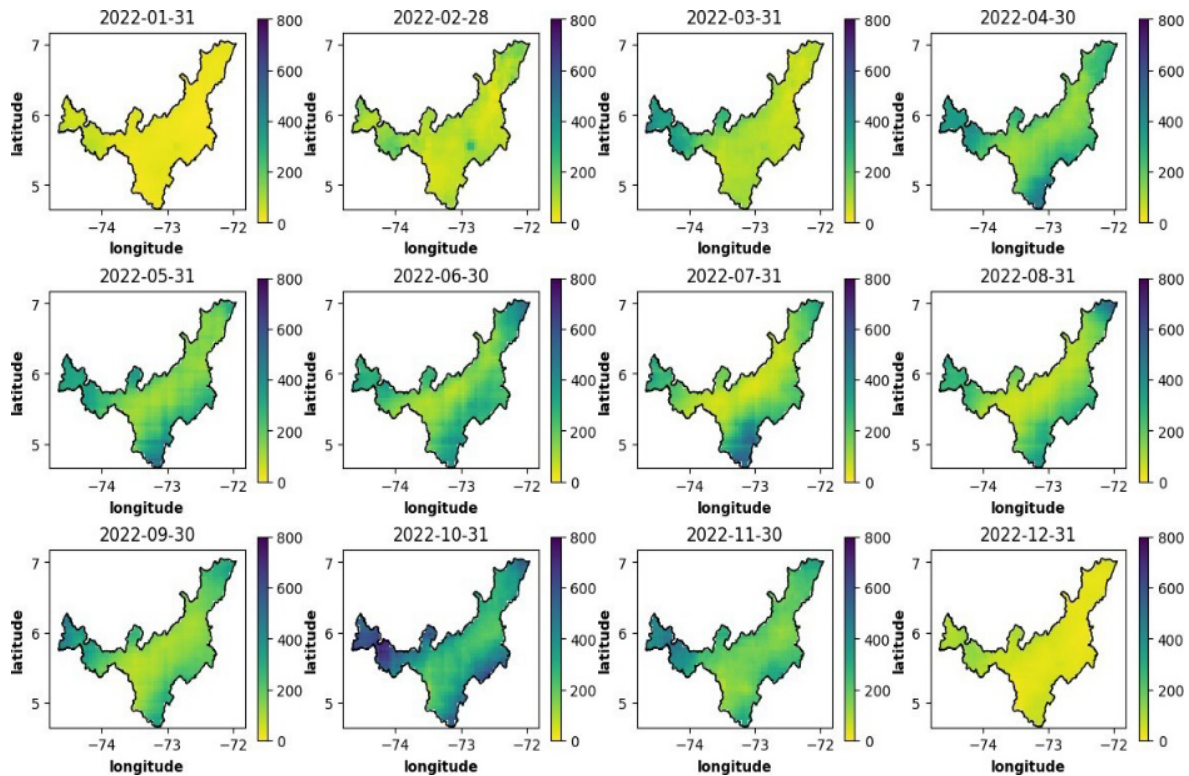


Figura 12. Precipitación mensual total por coordenada en el año 2022. Fuente: elaboración propia.

3.2 Evaluación de los modelos

Esta sección presenta una evaluación de los modelos utilizados en la investigación. Para comprender mejor el rendimiento de cada modelo, se han calculado y comparado métricas, como el MSE y la RMSE. Estas métricas son fundamentales para evaluar la capacidad de los modelos en la predicción de la precipitación mensual. La Tabla 2 presenta los resultados de las métricas de los modelos aplicados.

Tabla 2. Resultado de las métricas aplicadas. Fuente: elaboración propia.

	Montecarlo			DRL		
	Media	Mínimo	Máximo	Media	Mínimo	Máximo
MAE	34,363	7,789	103,291	68,627	12,480	240,206
MSE	2339,002	93,589	16089,823	11649,478	261,733	68397,7660
RMSE	43,624	9,674	126,845	80,0511	16,178	261,529

Se utilizaron varias métricas para evaluar el rendimiento de los modelos. En el modelo Montecarlo, el MAE promedio fue de 34,363, lo que significa que, en promedio, las predicciones se desviaron en aproximadamente 34,363 unidades de precipitación mensual de los valores reales. El MAE mínimo fue de 7,789, indicando precisión en algunos casos, mientras que el máximo fue de 103,291, mostrando desviaciones significativas en situaciones más desafiantes. Para el modelo DRL, el MAE promedio fue de 68,627, con un mínimo de 12,480 y un máximo de 240,206, lo que indica que las predicciones tuvieron desviaciones promedio de alrededor de 68,627 unidades de precipitación mensual, con casos de alta variabilidad.

El MSE se utilizó para cuantificar la magnitud promedio de los errores cuadráticos entre las predicciones y los valores reales. Para el modelo Montecarlo, el MSE promedio fue de 2339,002, con un mínimo de 93,589 y un máximo de 16089,823; reflejando errores cuadráticos de magnitudes variables. En el modelo DRL, el MSE promedio fue de 11649,478, con un mínimo de 261,733 y un máximo de 68397,7660, mostrando una magnitud moderada de errores cuadráticos con alta variabilidad.

El RMSE, que mide la magnitud de los errores en la misma unidad que los valores reales, indicó que en el modelo Montecarlo, el RMSE promedio fue de 43,624, con un mínimo de 9,674 y un máximo de 126,845. Para el modelo DRL, el RMSE promedio fue de 80,051, con un mínimo de 16,178 y un máximo de 261,529. Estas diferencias en la precisión pueden estar relacionadas con condiciones climáticas, ubicaciones geográficas o momentos específicos en Boyacá.

3.3 Comparación con los resultados de los modelos de RNNs y ConvLSTM

En esta sección se plantea la comparación de los resultados obtenidos por los modelos aplicados en este proyecto, específicamente el modelo de redes neuronales recurrentes con capas LSTM (RNNs con LSTM) y el modelo de redes neuronales convolucionales LSTM (ConvLSTM) comparados con los modelos de simulación Montecarlo y aprendizaje por refuerzo profundo (DRL). Para este propósito, se presentan las métricas clave en la Tabla 3, que incluyen el MSE, la RMSE y el coeficiente de determinación R².

Tabla 3. Resultado de las métricas de los modelos de RNNs y ConvLSTM. Fuente: elaboración propia.

	Modelo RNNs con LSTM	Modelo ConvLSTM
MSE	1401,37	3152,08
RMSE	37,43	56,14
R ²	0,876	0,722

Comparando los resultados de las métricas de MSE y RMSE entre los diferentes modelos, se destacan tendencias en su rendimiento. Los modelos analizados son: Montecarlo, Aprendizaje Profundo por Refuerzo (DRL), Redes Neuronales Recurrentes con capas LSTM (RRNSs con LSTM) y Redes Neuronales Convolucionales con LSTM (ConvLSTM).

En términos de MSE, que mide la magnitud promedio de los errores cuadráticos en las predicciones mensuales de precipitación, el modelo RRNSs con LSTM obtiene el valor más bajo, con un MSE de 1401,37, lo que indica una capacidad notable para reducir los errores en sus predicciones. Le sigue el modelo Montecarlo, con un MSE promedio de 2339,002, mostrando un buen rendimiento en la predicción de precipitación. Por otro lado, el modelo ConvLSTM registra un MSE más alto de 3152,08, lo que señala una mayor magnitud de errores en sus predicciones. Finalmente, el modelo DRL tiene el MSE más alto, con un promedio de 11649,478, reflejando una magnitud de errores significativamente mayor en sus predicciones.

El análisis del RMSE, que proporciona una medida más interpretable de la magnitud de los errores, respalda estas tendencias. Nuevamente, el modelo RRNSs con LSTM lidera con un RMSE de 37,43, indicando desviaciones promedio de 37,43 unidades de medida de precipitación en sus predicciones. Le sigue el modelo Montecarlo, con un RMSE promedio de 43,624, mostrando una capacidad razonable para predecir la precipitación. El modelo ConvLSTM, con un RMSE de 56,14, revela errores más significativos en sus predicciones. Por

último, el modelo DRL tiene el RMSE más alto de 80,0511, presentando la mayor magnitud de errores en las predicciones mensuales de precipitación.

Sin embargo, es fundamental recordar que las métricas MSE y RMSE, aunque útiles, no ofrecen una evaluación completa del rendimiento del modelo. La elección del modelo adecuado para una aplicación específica debe considerar otros factores como la complejidad computacional y la facilidad de implementación. En este caso, el modelo (DRL) resultó computacionalmente costoso y más difícil de implementar debido a la naturaleza del aprendizaje reforzado, lo que limitó su aplicabilidad en todas las predicciones.

3.4 Visualización de datos

Con el fin de ofrecer una representación visual enriquecedora y dinámica de los resultados obtenidos en el proyecto de predicción de precipitación mensual en el departamento de Boyacá, se ha creado un dashboard interactivo en Power BI. Este dashboard no solo proporciona una visión intuitiva de los datos reales históricos de precipitación, sino que también presenta las predicciones generadas por los modelos de simulación Montecarlo y aprendizaje reforzado profundo (DRL).

El dashboard se estructura en tres columnas, tal como se ilustra en la Figura 13, cada una enfocada en una categoría clave de información.

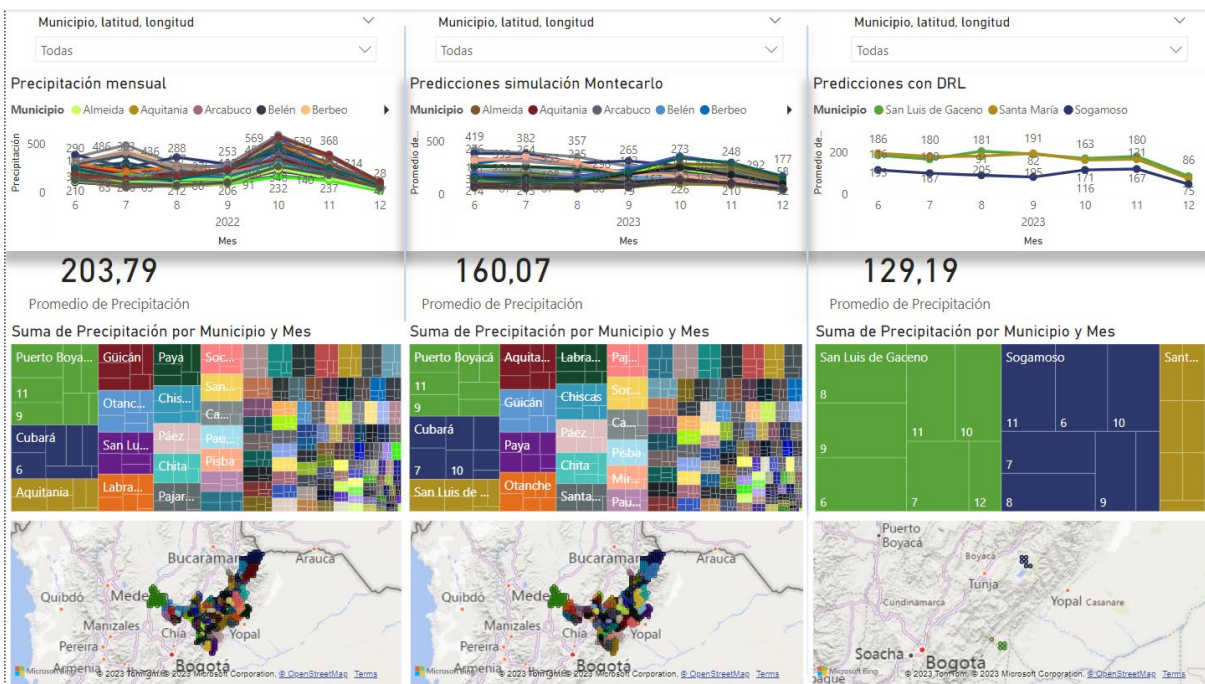


Figura 13. Estructura del Dashboard. Fuente: elaboración propia.

A continuación, una breve descripción de la estructura del dashboard.

- Histórico real de precipitación mensual: Muestra un gráfico de líneas con datos históricos de precipitación a lo largo de los años, permitiendo la comparación con las predicciones.
- Predicciones generadas por la simulación Montecarlo: Enfocado en las predicciones de este modelo, superpuestas con los datos históricos para evaluar su capacidad predictiva.

- Predicciones generadas por DRL: Presenta las predicciones generadas por el modelo DRL, comparándolas con los datos reales para determinar cuál modelo ofrece las predicciones más precisas.
- Treemap de precipitación por municipio y mes: Muestra la cantidad de precipitación en cada municipio durante diferentes meses, facilitando la identificación de patrones y distribuciones.
- Mapa georreferenciado: Visualiza la ubicación de 753 puntos en Boyacá, representando la cantidad de precipitación en cada punto. Esto proporciona una percepción espacial de las variaciones en la precipitación.

El filtrado geográfico, aplicado en este trabajo permitió delimitar el conjunto de datos a la zona geográfica específica seleccionada para el estudio, de manera que se pudo analizar la información de interés para conocer los valores máximos y mínimos de precipitación que se presentan en la región de Boyacá desde 1981 hasta 2023. Se observa una forma de visualización de información más detallada respecto a trabajos que solo incluyen los valores de precipitación que generan el fenómeno de la niña y el niño.

En este sentido, se llevó a cabo un análisis del municipio de Sogamoso empleando coordenadas geográficas precisas de latitud 5,63 y longitud -72,88, tal como se detalla en la Figura 14. Al realizar un examen de las series de tiempo y hacer referencia al gráfico correspondiente, se pudo constatar que ambos modelos utilizados en este proyecto siguen el patrón de precipitación en comparación con los datos históricos reales del año anterior.

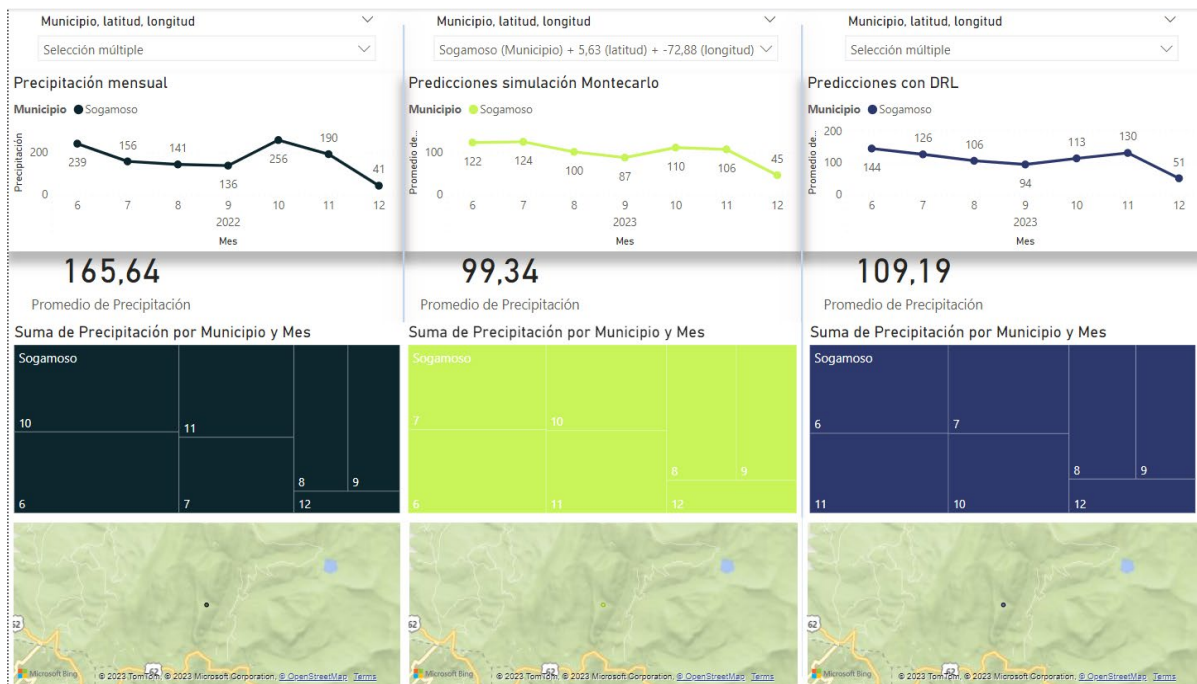


Figura 14. Visualización datos, municipio de Sogamoso. Fuente: elaboración propia.

Este análisis reveló que este modelo propuesto los meses con mayor registro de precipitación en esta ubicación de Sogamoso son junio, octubre y noviembre, mientras que los meses con la menor precipitación son septiembre y diciembre. Al analizar el promedio de los datos de precipitación pronosticados para los últimos seis meses del año 2023 se evidenció una disminución notable en ambos modelos [22]. Específicamente, el modelo de simulación

Montecarlo proyecta un promedio de 99,34 mm, mientras que el modelo DRL pronostica un promedio de 109,19 mm. Comparado con los valores históricos reales del año anterior [23], que presentaban un promedio de 165,64 mm, esto representa una disminución sustancial en la cantidad de lluvia esperada.

Adicionalmente, se realizó el análisis de manera similar a los anteriores. En esta ocasión, las coordenadas geográficas empleadas corresponden a la latitud 4,67 y longitud -73,18, asociadas al municipio de San Luis de Gaceno, tal como se exhibe en la Figura 15.

Respecto a modelo analizados y de acuerdo con los mapas espacio temporales de predicción de precipitación obtenidas por el modelo, para el año 2023 en el mes de septiembre será de poca lluvia, llegando a datos máximos de 100 mm y a partir de octubre hasta diciembre se tendrán lluvias moderadas en los municipios más cercanos a fronteras del departamento, presentando una ventaja significativa respecto a la exactitud de la predicción generada por otros algoritmos como [20].

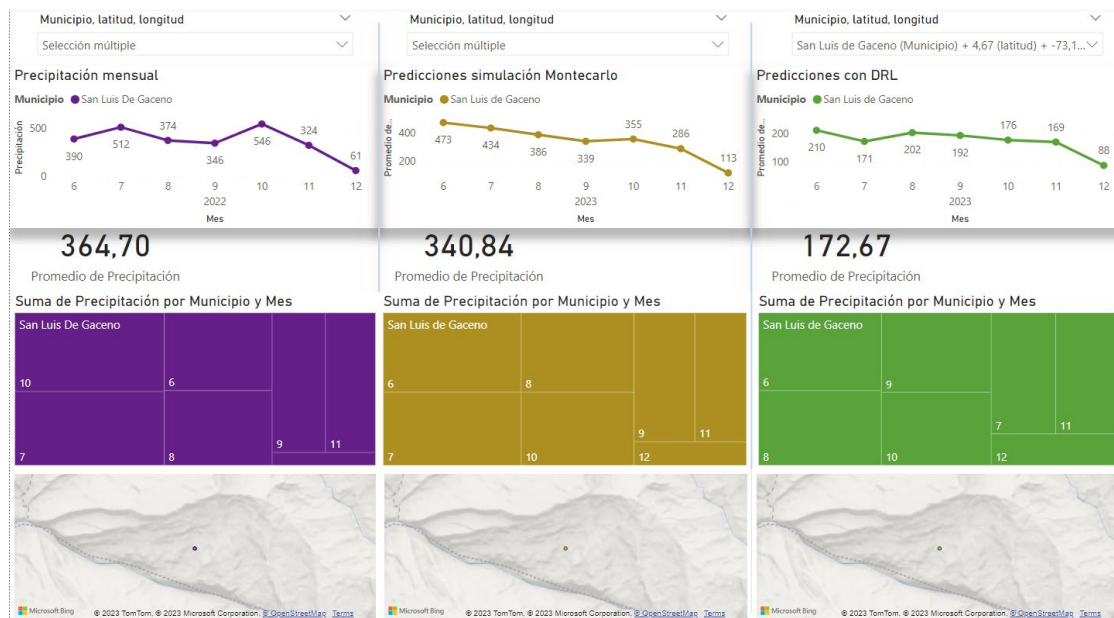


Figura 15. Visualización datos, municipio de San Luis de Gaceno. Fuente: elaboración propia.

Nuevamente, se destaca que los meses de junio y octubre son los de mayor registro de precipitación en esta ubicación, mientras que septiembre y diciembre representan los meses con menor precipitación. Al evaluar el promedio de los datos de precipitación pronosticados para los últimos seis meses de 2023, se observa una disminución considerable en ambos modelos. El modelo de simulación Montecarlo [13] proyecta un promedio de 340,84 mm, y el modelo DRL pronostica un promedio de 172,67 mm. En comparación con los valores históricos reales del año anterior, que promediaban 364,70 mm, esto implica una reducción significativa en la cantidad de lluvia esperada.

Se eligieron los municipios de Sogamoso y San Luis de Gaceno para los escenarios de visualización debido a dos razones clave. En primer lugar, las predicciones del modelo de aprendizaje reforzado profundo (DRL) [26] se limitaron a estos municipios debido a restricciones computacionales. En segundo lugar, se seleccionaron debido a su proximidad a importantes cuencas hidrográficas, como el río Chicamocha y la subcuenca del río Upía, que desemboca en el río Meta [27].

Para la implementación del modelo LSTM, se tuvo en cuenta el método de ventana usando 48 meses anteriores de precipitación para predecir 16 meses a partir de septiembre de 2023 [28], en donde la visualización de los resultados de predicción es fundamental para tomar decisiones o prevenciones en diferentes ámbitos. De tal manera que los resultados en mapas de calor en los diferentes puntos dentro del departamento de Boyacá permiten observar los patrones mensuales de la lluvia que varían durante el año a diferencia de modelos como [29], identificando los meses de precipitación más intensos (fenómeno de la niña), moderada y secos (fenómeno del niño) en los 123 municipios del departamento de Boyacá.

Las condiciones climáticas en estos municipios influyen directamente en el flujo de agua y la precipitación en estas cuencas, lo que impacta a las comunidades locales y actividades como la agricultura y la generación de energía hidroeléctrica. Al enfocarse en estos municipios, se facilita la comprensión de los patrones climáticos locales y su impacto en la gestión de recursos hídricos, contribuyendo a una toma de decisiones más informada en esta región.

4. CONCLUSIONES

Este proyecto, aplica la predicción de datos de precipitación mensual mediante el uso de modelos basados en aprendizaje reforzado (RL) y la integración de métodos de simulación de Montecarlo. En el proceso investigativo se llevó a cabo el desarrollo de dos enfoques distintos: El primer enfoque de investigación aplicada se basó en la aplicación de la simulación Montecarlo y el otro enfoque en investigación correlacional mediante la aplicación de aprendizaje por refuerzo. Se concluye que el aprendizaje por refuerzo no es la mejor alternativa para la predicción de variables cuantitativas continuas ya que en comparación con los modelos de redes neuronales se vio la gran diferencia entre las métricas de evaluación de los modelos, estando por encima de los resultados del modelo de aprendizaje reforzado (RL). Los resultados evidencian la capacidad predictiva de los modelos, proporcionando valiosas perspectivas sobre los patrones de precipitación en la región en estudio.

La implementación de la simulación Montecarlo demuestra ser una herramienta valiosa para generar predicciones precisas y coherentes con los datos reales. Esto brinda a interesados locales y a los sectores agrícola y de gestión de agua una ventaja significativa en la toma de decisiones basada en datos climáticos más precisos. El modelo DRL, aunque prometedor, presenta desafíos computacionales considerables debido a su intensidad de recursos. Esto limita su aplicabilidad en ciertas circunstancias. Además, se observa que la calidad de las predicciones puede variar según la cantidad de datos disponibles.

Se destaca la efectividad de los modelos de simulación Montecarlo y aprendizaje reforzado profundo (DRL) en la predicción de precipitación mensual en el departamento de Boyacá. Sin embargo, es importante contextualizar estas conclusiones en relación con otros modelos convencionales basados en Aprendizaje Profundo con Redes Neuronales, como LSTM o ConvLSTM.

Se ha demostrado que tanto la simulación Montecarlo como el DRL son capaces de generar predicciones significativas de la precipitación mensual en Boyacá. Estos modelos ofrecen una comprensión profunda de los patrones climáticos y permiten tomar decisiones informadas en varios sectores que dependen de datos climáticos precisos.

No obstante, es esencial reconocer que los modelos convencionales basados en Aprendizaje profundo, como LSTM o ConvLSTM, pueden superar a los enfoques de simulación Montecarlo y DRL en términos de precisión de predicción. Estos modelos, que se basan en redes neuronales profundas tienen la capacidad de capturar patrones temporales y

espaciales complejos en los datos climáticos; lo que a menudo se traduce en predicciones más precisas [30].

Aunque los modelos de Simulación Montecarlo y DRL ofrecen un enfoque innovador y efectivo para la predicción de precipitación mensual, no han superado la precisión y el rendimiento de los modelos convencionales basados en Aprendizaje Profundo con Redes Neuronales. Por lo tanto, se recomienda considerar estos últimos cuando la precisión en la predicción es de máxima importancia. Por otra parte, el desarrollo del tablero de control interactivo en Power BI ha demostrado ser altamente efectivo en el análisis de datos de precipitación en el departamento de Boyacá. Este tablero ofrece una representación visual dinámica de los datos climáticos y ha mejorado significativamente la comunicación de los resultados, así como la identificación de patrones y tendencias climáticas.

El desempeño del tablero se ha destacado gracias a su capacidad para proporcionar una comprensión detallada y accesible de los datos. Los mapas de calor geospaciales han permitido la visualización efectiva de la distribución espacial de la precipitación a lo largo del tiempo, lo que resulta esencial para identificar patrones en diferentes regiones de Boyacá y su evolución. Además, los gráficos de tendencias temporales han facilitado la comparación de predicciones con datos históricos, lo que es crucial para evaluar la precisión de los modelos en diferentes momentos.

El impacto de la topografía de Boyacá influye en la predicción de la precipitación, ya que el 24 % de las áreas de páramo de Colombia se encuentran en el departamento de Boyacá, por tanto, para identificar áreas montañosas, valles o llanuras, se sugiere un control de topografía incluidos en una variable o etiqueta para el entrenamiento del modelo y de esta manera mejorar la predicción y disminuir el error.

En cuanto a la idoneidad del caso de estudio, este tipo de dashboard interactivo resulta altamente adecuado para abordar los desafíos climáticos en la región. Boyacá, con su diversidad geográfica y variabilidad climática, se beneficia enormemente de la capacidad del tablero para representar datos geospaciales y mostrar tendencias temporales. La interacción intuitiva proporciona flexibilidad para adaptar el análisis a necesidades específicas, lo que es esencial en un campo en constante evolución como la climatología.

Para trabajos futuros, se puede considerar combinar modelos con la finalidad de disminuir los errores residuales y aumentar la exactitud o incluir la variable altitud, ya que como se indicó, la geografía en Colombia contiene unas zonas montañosas y otras planas, que pueden influir de manera significativa en precipitaciones intensas o bajas del modelo. Además, el resultado de este proyecto permite adecuar este modelo para generar predicciones a cualquier zona geográfica del País, simplemente conociendo la zona geográfica que se desea analizar.

5. AGRADECIMIENTOS Y FINANCIACIÓN

Los autores expresan su agradecimiento a la Vicerrectoría de investigación y Extensión de la UPTC; por el apoyo económico dado a través del proyecto SGI 3535.

CONFLICTOS DE INTERÉS DE LOS AUTORES

Los autores declaran que no se presentan conflictos de interés.

CONTRIBUCIÓN DE LOS AUTORES

Jimmy Alejandro Zea Gutiérrez: Programación de algoritmos de aprendizaje por refuerzo, pruebas al modelo y redacción del artículo.

Marco Javier Suárez Barón: Diseño y construcción de algoritmos de inteligencia artificial y redacción del artículo.

Juan Sebastián González Sanabria: Revisión metodológica.

6. REFERENCIAS

- [1] Esri. “Análisis y predicción de meteorología.” [arcgis.com](https://learn.arcgis.com/es/paths/meteorology/). Accessed: Feb. 19, 2024. [Online]. Available: <https://learn.arcgis.com/es/paths/meteorology/>
- [2] E. Mesgari, P. Mahmoudi, Y. Kord Tamandani, T. Tavousi, and S. M. Amir Jahanshahi, “A comparative assessment of the ability of different types of machine learning in short-term predictions of nocturnal frosts,” *Acta Geophysica*, vol. 72, no. 4, pp. 2955–2973, Aug. 2024. <https://doi.org/10.1007/S11600-023-01276-1>
- [3] IDEAM. “Predicción-IDEAM.” ideam.gov.co. Accessed: Feb. 19, 2024. [Online]. Available: <http://ideam.gov.co/web/siac/prediccion>
- [4] S. Zi-yi, and B. Wen-chao, “Machine learning model combined with CEEMDAN algorithm for monthly precipitation prediction,” *Earth Sci. Inform.*, vol. 16, no. 2, pp. 1821–1833, Jun. 2023. <https://doi.org/10.1007/S12145-023-01011-W>
- [5] A. Thamm GmbH. “Become a data.musketeer.” [alexanderthamm.com](https://www.alexanderthamm.com/de/karriere/). Accessed: Feb. 19, 2024. [Online]. Available: <https://www.alexanderthamm.com/de/karriere/>
- [6] M. Akbarian, B. Saghafian, and S. Golian, “Monthly streamflow forecasting by machine learning methods using dynamic weather prediction model outputs over Iran,” *Journal of Hydrology*, vol. 620, p. 129480, May. 2023. <https://doi.org/10.1016/J.JHYDROL.2023.129480>
- [7] J. Nithyashri, R. K. Poluru, S. Balakrishnan, M. Ashok Kumar, P. Prabu, and S. Nandhini, “IoT based prediction of rainfall forecast in coastal regions using deep reinforcement model,” *Measurement: Sensors*, vol. 29, p. 100877, Oct. 2023. <https://doi.org/10.1016/J.MEASEN.2023.100877>
- [8] Y. Jun-He, C. Ching-Hsue, and C. Chia-Pan, “A time-series water level forecasting model based on imputation and variable selection method,” *Comput. Intell. Neurosci.*, vol. 2017, pp. 1-11, Nov. 2017. <https://doi.org/10.1155/2017/8734214>
- [9] R. Yang, H. Liu, N. Nikitas, Z. Duan, Y. Li, and Y. Li, “Short-term wind speed forecasting using deep reinforcement learning with improved multiple error correction approach,” *Energy*, vol. 239, p. 122128, Jan. 2022. <https://doi.org/10.1016/J.ENERGY.2021.122128>
- [10] M. Zhang, Z. Xu, Y. Wang, S. Zeng, and X. Dong, “Evaluation of uncertain signals’ impact on deep reinforcement learning-based real-time control strategy of urban drainage systems,” *J. Environ. Manage.*, vol. 324, p. 116448, Dec. 2022. <https://doi.org/10.1016/J.JENVMAN.2022.116448>
- [11] K. Skarlatos, E. S. Bekri, D. Georgakellos, P. Economou, and S. Bersimis, “Projecting Annual Rainfall Timeseries Using Machine Learning Techniques,” *Energies*, vol. 16, no. 3, p. 1459, Feb. 2023, <https://doi.org/10.3390/EN16031459>
- [12] P. D. Nolasco Ramírez, “Aplicación de Machine Learning para pronóstico de desplazamiento de lluvias usando imágenes del radar de lluvias de UDEP,” Trabajo de Pregrado, Universidad de Piura, Piura, Perú, 2023. [Online]. Available: <https://hdl.handle.net/11042/6007>
- [13] R. He, L. Zhang, and A. W. Z. Chew, “Data-driven multi-step prediction and analysis of monthly rainfall using explainable deep learning,” *Expert Systems with Applications*, vol. 235, p. 121160, Jan. 2024. <https://doi.org/10.1016/J.ESWA.2023.121160>
- [14] A. U. G. Senocak, M. T. Yilmaz, S. Kalkan, I. Yucel, and M. Amjad, “An explainable two-stage machine learning approach for precipitation forecast,” *Journal of Hydrology*, vol. 627, p. 130375, Dec. 2023. <https://doi.org/10.1016/J.JHYDROL.2023.130375>
- [15] C. Wang, J. Xu, G. Tang, Y. Yang, and Y. Hong, “Infrared Precipitation Estimation Using Convolutional Neural Network,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8612–8625, Dec. 2020. <https://doi.org/10.1109/TGRS.2020.2989183>
- [16] Q. Yuan *et al.*, “Deep learning in environmental remote sensing: Achievements and challenges” *Remote Sensing of Environment*, vol. 241, p. 111716, 2020. <https://doi.org/10.1016/j.rse.2020.111716>

- [17] T. Nan *et al.*, “Evaluation of shallow groundwater dynamics after water supplement in North China Plain based on attention-GRU model,” *Journal of Hydrology*, vol. 625, p. 130085, Oct. 2023. <https://doi.org/10.1016/J.JHYDROL.2023.130085>
- [18] E. E. León Baque, V. D. Vásquez Granda, and M. D. Valderrama Chávez, “Cambios en patrones de precipitación y temperatura en el Ecuador: regiones sierra y oriente,” *Dilemas contemporáneos: educación, política y valores*, vol. 8, no. SPE, Mar. 2021. <https://doi.org/10.46377/DILEMAS.V8I.2608>
- [19] C. J. Fernández Pérez, G. E. Cely Reyes, and P. A. Ramírez, “Cuantificación de la captura de carbono y análisis de las propiedades del suelo en coberturas naturales y una plantación de pino en el páramo de Rabanal, Colombia,” *Cuad. Geogr. Rev. Colomb. Geogr.*, vol. 28, no. 1, pp. 121–133, Jan. 2019. <https://doi.org/10.15446/RCDG.V28N1.66152>
- [20] E. Morales Rojas, E. A. Díaz Ortiz, L. García, and M. E. Milla Pino, “Pronóstico de precipitaciones mensuales: Un estudio de caso en las comunidades nativas del Perú,” *Revista Científica Pakamuros*, vol. 9, no. 3, pp. 71–85, Sep. 2021. <https://doi.org/10.37787/13azmg02>
- [21] D. M. Herrera Posada, and E. Aristizábal, “Modelo de inteligencia artificial y aprendizaje automático para la predicción espacial y temporal de eventos de sequía en el departamento del Magdalena, Colombia,” *INGE CUC*, vol. 18, no. 2, pp. 249–265, Nov. 2022. <https://doi.org/10.17981/INGECUC.18.2.2022.20>
- [22] E. Duarte. “Rainfall data for the department of Boyacá.” kaggle.com. Accessed: Feb. 19, 2024. [Online]. Available: <https://www.kaggle.com/datasets/estebanduarte/rainfall-data-for-the-department-of-boyaca>
- [23] M. A. Saleh, and H. M. Rasel, “Performance evaluation of Machine Learning based regression models for rainfall forecasting,” *Research Square*, vol. 25, Jan. 2024. <https://doi.org/10.21203/RS.3.RS-3856741/V1>
- [24] S. D. Latif, and A. N. Ahmed, “A review of deep learning and machine learning techniques for hydrological inflow forecasting,” *Environment, Development and Sustainability*, vol. 25, no. 11, pp. 12189–12216, Mar. 2023. <https://doi.org/10.1007/S10668-023-03131-1>
- [25] S. Eddamiri, F. Z. Bassine, V. Ongoma, T. Epule Epule, and A. Chehbouni, “An automatic ensemble machine learning for wheat yield prediction in Africa,” *Multimed. Tools Appl.*, Jan. 2024. <https://doi.org/10.1007/S11042-024-18142-X>
- [26] D. Elavarasan, and P. M. Durairaj Vincent, “Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications,” *IEEE Access*, vol. 8, pp. 86886–86901, May. 2020. <https://doi.org/10.1109/ACCESS.2020.2992480>
- [27] J. B. Valencia, V. V. Guryanov, J. Mesa-Diez, N. Diaz, D. Escobar-Carbonari, and A. V. Gusarov, “Predictive Assessment of Climate Change Impact on Water Yield in the Meta River Basin, Colombia: An InVEST Model Application,” *Hydrology*, vol. 11, no. 2, p. 25, Feb. 2024. <https://doi.org/10.3390/HYDROLOGY11020025>
- [28] G. Sharma, A. Singh, and S. Jain, “A hybrid deep neural network approach to estimate reference evapotranspiration using limited climate data,” *Neural Comput. Appl.*, vol. 34, no. 5, pp. 4013–4032, Mar. 2022. <https://doi.org/10.1007/S00521-021-06661-9>
- [29] S. Duarte, G. A. Corzo Perez, G. Santos, and D. P. Solomatine, “Application of Natural Language Processing to Identify Extreme Hydrometeorological Events in Digital News Media: Case of the Magdalena River Basin, Colombia,” in *Advanced Hydroinformatics: Machine Learning and Optimization for Water Resources*, G. A. Corzo Perez, D. P. Solomatine, Ed., New York, NY, USA: Wiley, 2024, pp. 283–318. <https://doi.org/10.1002/9781119639268.CH10>
- [30] Z. Gao *et al.*, “PreDiff: Precipitation Nowcasting with Latent Diffusion Models,” 2023, *arXiv: 2307.10422*. <https://arxiv.org/abs/2307.10422v2>