

Bogotá D.C., 21 de abril de 2025

REPORTE FASE 1

TRABAJO FINAL DE MAESTRÍA “PRONÓSTICO DE LA PRECIPITACIÓN ACUMULADA EN COLOMBIA, UTILIZANDO TÉCNICAS DE APRENDIZAJE AUTOMÁTICO”

NOMBRE: JUAN DIEGO MOGOLLÓN OVIEDO	
OBJETIVO: Desarrollar un modelo para el pronóstico de la precipitación acumulada en Colombia, utilizando técnicas de aprendizaje automático.	
OBJETIVOS ESPECÍFICOS: <ul style="list-style-type: none">• Seleccionar un conjunto de datos abiertos de precipitación acumulada en Colombia, verificando su acceso y calidad mediante algoritmos de análisis de datos.• Implementar un proceso para la extracción, transformación y carga de los datos, aplicando técnicas de ingeniería de características.• Evaluar el desempeño de diferentes algoritmos de aprendizaje automático para el pronóstico de precipitación acumulada, utilizando los datos procesados y métricas de evaluación empleadas en predicción climática.• Diseñar una arquitectura para la implementación del modelo en un entorno de prueba, aplicando prácticas y herramientas de operaciones de aprendizaje automático.	
PLAZO DE EJECUCIÓN: 16 semanas	
INICIO: 31 de marzo de 2025	FINALIZACIÓN: 26 de julio de 2025

Con el objetivo de dar cumplimiento al cronograma estipulado en la “Propuesta de Trabajo Final de Maestría” en el periodo 31 de marzo de 2025 al 20 de abril de 2025, se presenta a continuación el reporte de la fase 1:

<p>RESULTADO OBTENIDO: Se seleccionó el conjunto de datos abiertos de precipitación acumulada diaria de las estaciones meteorológicas del Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM, verificando su acceso y calidad por medio de algoritmos de análisis de datos.</p> <p>Se obtuvieron un total de 14.762.874 de registros de precipitación acumulada diaria en el periodo 2000 – 2024, de 2036 estaciones meteorológicas distribuidas en varias zonas del país. Se identificaron 225 estaciones con un gran porcentaje de datos faltantes, 273 estaciones con persistencia de ceros, 4 estaciones con persistencia de lluvia, patrones trimestrales y una mayor concentración de los datos entre 0 - 5 mm. A su vez, se encontró una correlación moderada a alta en los patrones de 7 días de lluvia de todas las estaciones.</p> <p>A continuación, se presentan las gráficas que soportan los resultados mencionados anteriormente.</p>

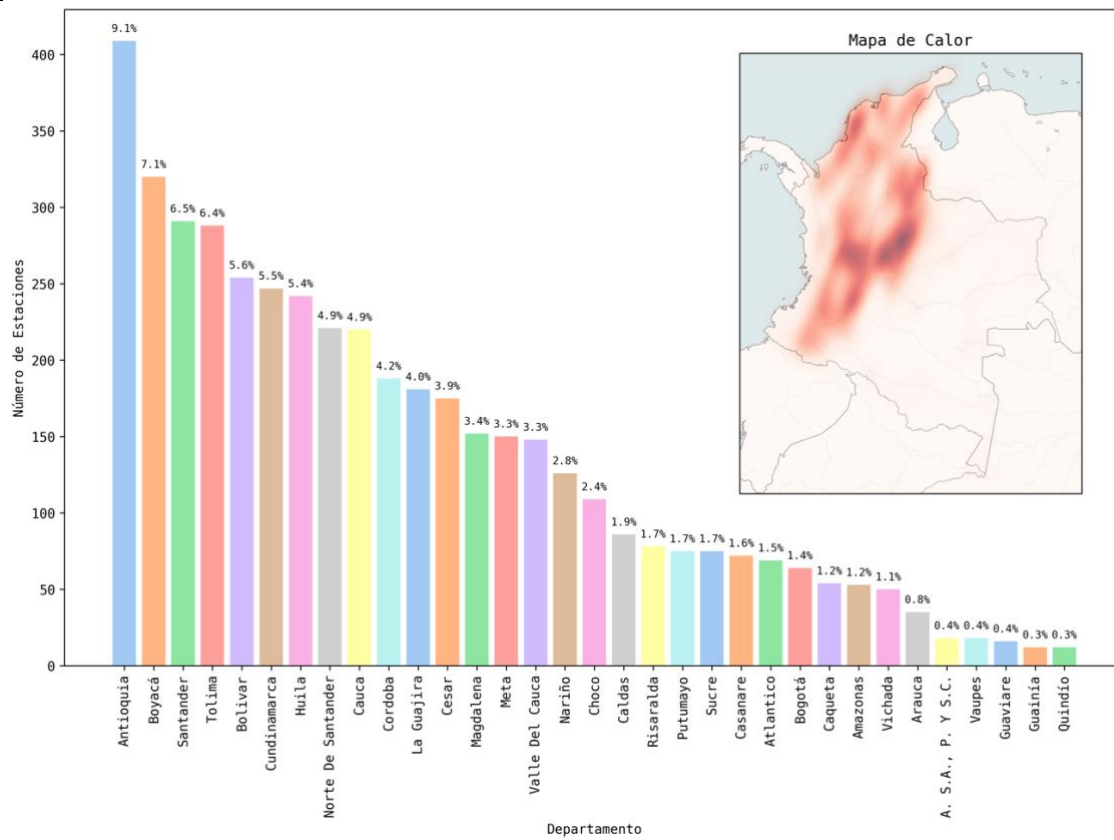


Figura 1. Distribución de las estaciones meteorológicas por departamento.

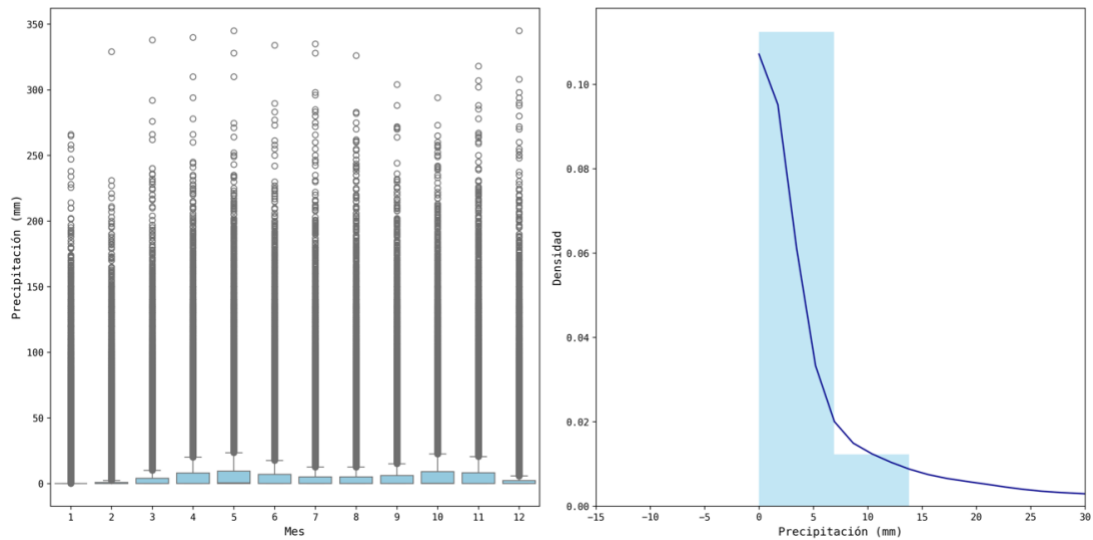


Fig. 2. Izquierda: Boxplot de la precipitación acumulada diaria por mes. Derecha: Distribución de densidad de precipitación acumulada diaria.

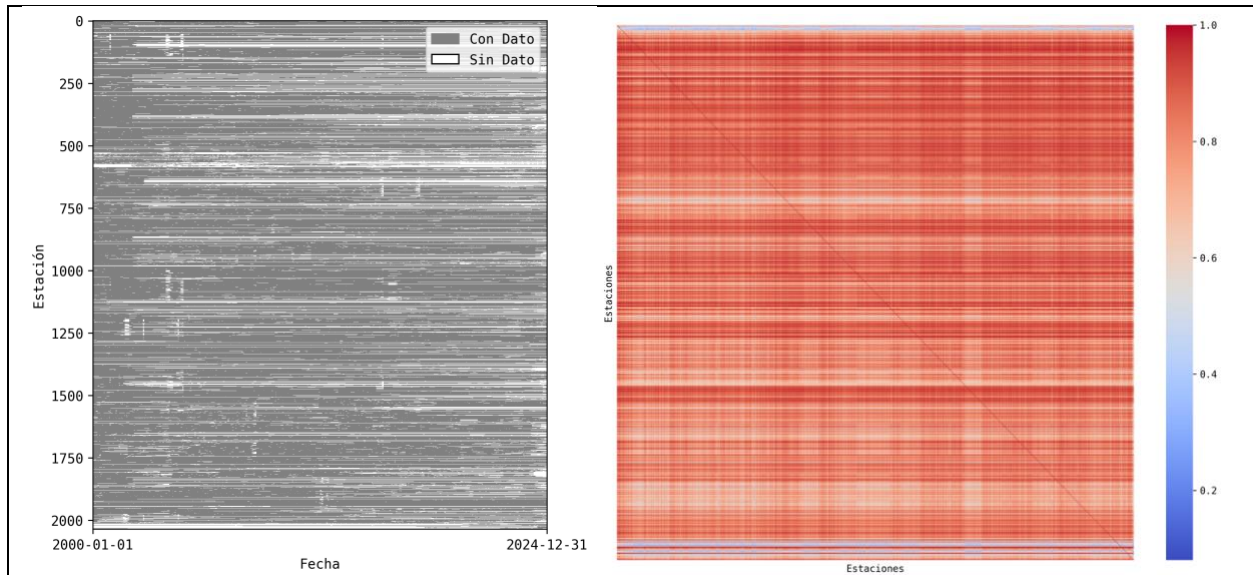


Fig. 3. Izquierda: Presencia o ausencia de datos de precipitación acumulada diaria en el periodo 2000 - 2024. Derecha: Matriz de correlación de patrones de 7 días de precipitación acumulada diaria de todas las estaciones.

ACTIVIDADES REALIZADAS:

1. Se estableció el contexto del proyecto, definiendo los antecedentes, el alcance y el plan a seguir.
2. Se definió la arquitectura inicial del sistema, sus restricciones y escalabilidad. (Esta actividad se complementará en la fase 4)
3. Se identificaron métodos y algoritmos de ML aplicables al pronóstico de precipitación acumulada y sus requerimientos en el contexto del proyecto.
4. Se identificaron los datos necesarios de acuerdo con los requerimientos establecidos.
5. Se realizó una búsqueda de la fuente de los datos para establecer el enlace de conexión a la API de la plataforma DHIME del IDEAM, utilizando la herramienta inspeccionar del navegador. Con ello, se desarrolló el script “data_load.ipynb” para la descarga de los datos de precipitación acumulada diaria de las estaciones meteorológicas.
6. Se desarrolló el script “eda.ipynb” con el fin de realizar un análisis exploratorio de los datos.

ENLACE: [Repositorio - GitHub](#)

Jorge Eliecer Camargo Mendoza
Director

Juan Diego Mogollón Oviedo
Estudiante