

PREDICCIÓN DEL RECURSO HÍDRICO EN LAS CUENCAS DEL MUNICIPIO DE MANIZALES



Participantes

Luisa Fernanda Pachón Sánchez
Daniel Aguilar Tenjo
Jhon Fredy Vallejo Idárraga
Daniel Perdomo Sánchez
Miguel Mejía Arango

Ejecutores

Margarita Maria Orozco
Natalia Betancur Herrera
Frank Yesid Zapata Cataño

Tabla de contenido

1.	Introducción	3
2.	Planteamiento del problema	4
3.	Objetivos	4
3.1.	Objetivo General	4
3.2.	Objetivos específicos	4
4.	Justificación.....	4
5.	Alcance	5
6.	Presupuesto	5
6.1.	Costos Directos	5
6.1.1.	Recursos Humanos	5
6.1.2.	Equipos y Herramientas	6
6.1.3.	Trabajo de Campo.....	6
6.1.4.	Infraestructura y Logística	6
6.2.	Costos Indirectos	7
6.2.1.	Administración y Gestión del Proyecto.....	7
6.2.2.	Imprevistos y Contingencias	7
6.3.	Resumen de Costos	7
7.	Metodología CRISP-DM	7
7.1.	Entendimiento del Negocio.....	8
7.2.	Entendimiento de los Datos	8
7.3.	Preparación de los Datos	8
7.4.	Modelado.....	8
7.5.	Evaluación	9
7.6.	Despliegue.....	9
7.7.	Conclusiones.....	9
8.	Modelado de los datos	9
8.1.	Características de los datos.....	9
8.2.	Tipo de datos:.....	10
8.3.	Características principales:	10

8.4.	Observaciones:	11
8.5.	Correlación de los datos	12
8.5.1.	Matriz de correlación #1: Sin día y Mes y Sin Codificar.	12
8.5.2.	Matriz de correlación # 2 día y mes codificados con variable velocidad	13
8.5.3.	Matriz de correlación #3: día y mes sin codificar y sin variable velocidad.....	15
8.5.4.	Matriz de correlación #4 Día y Mes codificadas y sin variable velocidad.	16
8.6.	Resultado de correlación.....	17
9.	Resultados del Modelo	18
10.	Anexos.....	21
11.	Bibliografía.....	21

1. Introducción

El cambio climático es uno de los desafíos globales más apremiantes del siglo XXI, afectando numerosos aspectos del medio ambiente y las sociedades humanas. Uno de los recursos más vulnerables y críticos en este contexto es el agua. El cambio climático altera los patrones meteorológicos y provoca fenómenos como el aumento de las temperaturas, cambios en la frecuencia e intensidad de las precipitaciones, derretimiento de los glaciares y eventos climáticos extremos, como sequías e inundaciones. Estos impactos están modificando la disponibilidad, distribución y calidad del recurso hídrico en todo el mundo.

A nivel global, el cambio climático está afectando tanto la oferta como la demanda de agua. En regiones donde históricamente el recurso ha sido abundante, como ciertas áreas de América del Norte y Europa, la disponibilidad de agua está disminuyendo debido a patrones de precipitación alterados y al estrés en las fuentes subterráneas. En contraste, regiones áridas y semiáridas, como el norte de África y partes de Asia y América Latina, se enfrentan a una intensificación de la escasez hídrica. La sobreexplotación de acuíferos, el uso insostenible de los recursos hídricos y el aumento de la demanda debido al crecimiento poblacional y económico agravan aún más esta situación.

En resumen, el cambio climático ha generado un impacto directo en la seguridad hídrica, especialmente en países como Colombia, caracterizado por su biodiversidad y variabilidad climática, la situación es especialmente crítica, el aumento en la frecuencia de fenómenos climáticos extremos, como sequías e inundaciones, ha afectado tanto a la cantidad como la calidad del agua, exacerbando la presión sobre este recurso vital.

Las regiones montañosas como la zona andina, donde se encuentra ubicado el Municipio de Manizales, dependen de cuencas hidrográficas sensibles a los cambios en el clima, así como el rápido crecimiento urbano y la expansión agrícola han incrementado la presión sobre estos recursos, poniendo en riesgo la sostenibilidad hídrica a largo plazo.

Por lo tanto, la gestión sostenible del agua se está volviendo cada vez más crítica en un mundo que enfrenta los efectos del cambio climático, la integración de herramientas científicas y tecnológicas avanzadas, como la inteligencia artificial y los modelos predictivos, son esenciales para anticipar las fluctuaciones en la disponibilidad de agua y desarrollar estrategias efectivas para la adaptación y mitigación de los impactos.

2. Planteamiento del problema

El Municipio de Manizales enfrenta serios desafíos en la gestión de sus recursos hídricos debido a la alta variabilidad climática, el crecimiento urbano y la deforestación, fenómenos como El Niño y La Niña alteran los patrones de precipitación, afectando la disponibilidad de agua, mientras que el aumento de la población y la expansión urbana incrementa la demanda sobre fuentes hídricas.

La deforestación en las cuencas disminuye su capacidad de regulación natural, y la infraestructura actual no siempre puede enfrentar eventos climáticos extremos. Además, la falta de modelos avanzados para predecir la disponibilidad de agua limita la capacidad de planificar a largo plazo, lo que genera riesgos de desabastecimiento, vulnerabilidad ante desastres y problemas socioeconómicos. Ante esta situación, se propone desarrollar un modelo de inteligencia artificial que permita predecir la disponibilidad hídrica en las cuencas del municipio, integrando variables climáticas, geográficas y de uso del suelo, con el fin de mejorar la toma de decisiones y garantizar una gestión sostenible del agua.

3. Objetivos

3.1. Objetivo General

Desarrollar un modelo de inteligencia artificial para la predicción de los recursos hídricos en las cuencas del Municipio de Manizales, con el fin de optimizar la gestión sostenible del agua y apoyar la toma de decisiones en su planificación y administración.

3.2. Objetivos específicos

- 3.2.1. Recolectar y analizar datos climáticos e hidrológicos históricos de las cuencas del municipio de Manizales para identificar patrones claves que influyen en la disponibilidad de recursos hídricos y preparar los datos para su uso en el desarrollo de un modelo de predicción.
- 3.2.2. Desarrollar y validar un modelo de predicción basado en técnicas de aprendizaje automático, que utilice variables climáticas e hidrológicas.

4. Justificación

El proyecto Predicción del recurso hídricos en las cuencas del Municipio de Manizales busca aportar una solución innovadora y basada en inteligencia artificial

(IA) a uno de los mayores desafíos que enfrenta tanto la región como el país La gestión eficiente y sostenible del recurso hídrico, a través del desarrollo de un modelo predictivo, los entes gubernamentales podrán tomar decisiones más informadas y planificar políticas que aseguren la disponibilidad del recurso hídrico a corto, mediano y largo plazo. Esto permitirá optimizar la asignación de recursos, anticipar períodos de escasez o abundancia de agua, y mejorar la respuesta frente a eventos climáticos extremos y reduciendo el riesgo de desastres naturales.

Para la comunidad, este proyecto será crucial para garantizar el acceso al agua potable, promoviendo su uso sostenible en actividades agrícolas, industriales y domésticas. Además, impulsará el desarrollo de proyectos de energía renovable, en especial hidroeléctricos, al ofrecer pronósticos más precisos sobre la disponibilidad de agua, clave para la viabilidad de este tipo de iniciativas.

Los investigadores y desarrolladores del proyecto encontrarán una oportunidad para aplicar técnicas avanzadas de IA, contribuyendo al avance del conocimiento científico y técnico en Colombia.

A nivel nacional, este proyecto será un referente en la planificación hídrica, que podrá replicarse en otras cuencas del país, fortaleciendo la resiliencia de Colombia frente al cambio climático y consolidando su liderazgo en el uso de tecnologías de punta para la gestión de recursos naturales.

5. Alcance

El proyecto se centrará en las cuencas hidrográficas dentro de los límites del municipio de Manizales, que son fundamentales para el abastecimiento de agua potable, la actividad agrícola y la conservación del ecosistema.

6. Presupuesto

6.1. Costos Directos

6.1.1. Recursos Humanos

Incluye el pago de salarios o honorarios de las personas directamente involucradas en la ejecución del proyecto.

Cargo	Cantidad	Costo Mensual	Duración	Total
<i>Investigadores</i>	5	0	2 meses	0

Cargo	Cantidad	Costo Mensual	Duración	Total
<i>Científico de Datos</i>	<i>2</i>	<i>7.000.000</i>	<i>2 meses</i>	<i>14.000.000</i>
<i>Ingeniero Ambiental</i>	<i>1</i>	<i>4.500.000</i>	<i>1 meses</i>	<i>4.500.000</i>
<i>Programador (IA y Modelos)</i>	<i>1</i>	<i>5.000.000</i>	<i>1 meses</i>	<i>5.000.000</i>

Subtotal Recursos Humanos: \$23.500.000

6.1.2. Equipos y Herramientas

Inversión en hardware y tecnología necesaria para procesar y almacenar los datos.

Ítem	Cantidad	Costo unitario	Total
<i>Estaciones de trabajo (PC)</i>	<i>3</i>	<i>0</i>	<i>0</i>

Subtotal Equipos y Herramientas: \$0

6.1.3. Trabajo de Campo

Incluye costos de desplazamientos para la recolección de datos en las cuencas seleccionadas.

Ítem	Cantidad	Costo unitario	Total
<i>Vehículos y combustible</i>	<i>2 meses</i>	<i>200.000</i>	<i>400.000</i>

Subtotal Trabajo de Campo: \$400.000

6.1.4. Infraestructura y Logística

Costos asociados a la gestión de espacios físicos y suministros necesarios.

Ítem	Cantidad	Costo Mensual	Duración	Total
<i>Suministros de papelería e insumos</i>	<i>N/A</i>	<i>50.000</i>	<i>2 meses</i>	<i>100.000</i>

Subtotal Infraestructura y Logística: \$100.000

6.2. Costos Indirectos

6.2.1. Administración y Gestión del Proyecto

Concepto	Cantidad	Costo Mensual	Costo
Conectividad (planes de voz y datos)	7	43.000	301.000
Servicios públicos (Agua, luz, internet)	1	300.000	600.000

Subtotal Administración: \$901.000

6.2.2. Imprevistos y Contingencias

Reserva para imprevistos que puedan surgir durante la ejecución del proyecto (10% de los costos directos).

Concepto	Porcentaje	Costo
Imprevistos y contingencias	10%	\$2.492.490

6.3. Resumen de Costos

Categoría	Total
Recursos Humanos	\$23.500.000
Equipos y Herramientas	\$0
Trabajo de Campo	\$400.000
Infraestructura y Logística	\$100.000
Administración y Gestión del Proyecto	\$901.000
Imprevistos y Contingencias	\$2.492.490
Total del Proyecto	\$27.417.390

7. Metodología CRISP-DM

Para el desarrollo del proyecto de predicción de recursos hídricos en las cuencas del Municipio de Manizales, seguimos la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) estándar internacional ampliamente utilizado en la minería de datos y ciencia de datos, que organiza el proceso en seis fases

secuenciales, las cuales nos permitieron estructurar y ejecutar el proyecto de manera eficiente como se describe a continuación:

7.1. Entendimiento del Negocio

Inicialmente, se llevó a cabo una serie de reuniones para comprender los desafíos asociados con la disponibilidad de recursos hídricos en la región. Definimos como objetivo principal el desarrollo de un modelo predictivo que pudiera anticipar la disponibilidad de agua en diferentes cuencas, ayudando a optimizar la gestión y apoyar la planificación sostenible del agua. Establecimos métricas clave para evaluar el éxito del proyecto, como la precisión en la predicción de caudales y la capacidad de identificar períodos de riesgo de escasez.

7.2. Entendimiento de los Datos

Recopilamos datos históricos sobre caudales de ríos, precipitaciones, temperaturas, humedad y otros factores climáticos de diversas fuentes, incluyendo estaciones meteorológicas locales y bases de datos regionales. Se realizó un análisis exhaustivo de los datos para identificar patrones, tendencias estacionales y posibles problemas de calidad, como valores atípicos y datos faltantes. Además, se analizaron correlaciones entre las variables para comprender qué factores tenían mayor influencia en la disponibilidad de recursos hídricos.

7.3. Preparación de los Datos

Los datos recopilados fueron limpiados para eliminar inconsistencias. Se seleccionaron las variables más relevantes y se crearon nuevas características que mejoraron la capacidad predictiva del modelo, como índices de sequía y acumulaciones de precipitación a lo largo del tiempo. Se aplicaron técnicas de normalización para estandarizar los datos, asegurando que fueran adecuados para los algoritmos de modelado.

7.4. Modelado

Probamos múltiples enfoques de modelado para determinar cuál proporcionaba las mejores predicciones. Después de varias rondas de ajuste y evaluación, seleccionamos un modelo basado en técnicas de aprendizaje automático que ofrecía la mejor combinación de precisión y robustez frente a datos de entrada variados. El modelo final fue ajustado con técnicas de validación cruzada para garantizar que las predicciones fueran precisas y consistentes, incluso para datos que no se habían visto antes.

7.5. Evaluación

Evaluamos el rendimiento del modelo utilizando métricas como el error cuadrático medio (MSE) y el coeficiente de determinación (R^2), que mostraron altos niveles de precisión en la predicción de los recursos hídricos. Además, se realizaron pruebas con datos de periodos recientes para validar que el modelo pudiera anticipar correctamente la disponibilidad de agua en diferentes condiciones climáticas. Los resultados fueron revisados por expertos, quienes confirmaron que el modelo proporcionaba información valiosa para la planificación de la gestión del agua.

7.6. Despliegue

Implementamos el modelo en un sistema de predicción en tiempo real que se integra con las bases de datos de monitoreo climático de Manizales. Este sistema permite la actualización continua del modelo con nuevos datos, garantizando que las predicciones se mantengan actualizadas. Se desarrolló una interfaz accesible para que los tomadores de decisiones puedan visualizar las predicciones y recibir alertas tempranas sobre posibles periodos de escasez. Además, se capacitó al personal en el uso del sistema y en la interpretación de las predicciones, asegurando que puedan tomar decisiones informadas basadas en los resultados del modelo.

7.7. Conclusiones

La implementación de la metodología CRISP-DM permitió que el proyecto se ejecutara de manera estructurada y eficiente, asegurando que cada fase fuera cuidadosamente planeada y ajustada para cumplir con los objetivos establecidos. Gracias a este enfoque, logramos desarrollar un modelo predictivo preciso y robusto que apoya la gestión sostenible del agua en Manizales, proporcionando información clave para la planificación y administración de los recursos hídricos.

8. Modelado de los datos

8.1. Características de los datos

A continuación, se realiza el análisis y descripción de la estructura, calidad y distribución de los datos, se describe el tipo y cantidad de variables y se calculan estadísticas básicas (promedios, medianas, valores extremos), se identifican valores faltantes, duplicados y atípicos, además de estudiar la distribución de las variables y posibles correlaciones o asociaciones entre ellas, así mismo, se verifica la

consistencia de los datos y se documentan las decisiones y hallazgos, lo cual permite preparar los datos para el desarrollo del modelo predictivo.

Filas sin duplicados: La base de datos contiene registros únicos, lo que significa que no hay filas repetidas. Esto es importante para mantener la integridad de los datos y asegurar que cada entrada sea distinta.

Tamaño de la base de datos: Con un tamaño de 294.8 KB en RAM, la base de datos es relativamente pequeña, lo que facilita su manejo y procesamiento.

Variables: La base de datos tiene un total de 8 variables, que representan diferentes aspectos de los datos que se están analizando tales como código estación, Nombre estación, Fecha, Nivel, Precipitación, Temperatura, Humedad y Velocidad.

8.2. Tipo de datos:

Categorico: Estas características representan categorías o grupos. Pueden ser variables cualitativas que no tienen un orden inherente, en la base se encuentran 2 variables categóricas que son: Nombre de estación y fecha, ambas columnas categóricas están completas (sin valores faltantes), sin embargo, cada una tiene solo un valor único, indicando que se refiere a una única estación en el análisis.

Numérico: Estas son variables cuantitativas que representan valores numéricos, en la base se encuentran 5 variables numéricas que son: Nivel, precipitación, temperatura, humedad y velocidad.

8.3. Características principales:

Código Estación: Todos los valores están presentes (100% de las filas), sin embargo, solo hay un valor distinto, lo que significa que todos los datos provienen de la misma estación.

Nombre Estación: Todos los valores están presentes (100% de las filas) sin embargo, contiene un único valor distinto, por lo que no aporta variación.

Fecha: Contiene 2,183 valores distintos, lo que indica una frecuencia de recopilación diaria, las fechas varían desde el 1 de febrero de 2016 hasta el 29 de febrero de 2020. Todos los valores están presentes, y no hay valores faltantes, cada fila tiene un valor único, lo que sugiere que los datos están bien diferenciados.

Nivel: No cuenta con valores faltantes, el rango de los valores va de 110 a 206.3, con una media de 132.6, hay cierto sesgo positivo (1.21) y una leve curtosis (3.01) es decir, la mayoría de los datos se agrupan en valores más bajos, pero hay algunos valores inusualmente altos que influyen la distribución hacia la derecha.

Precipitación: El 48% de los datos están ausentes, un 21% de los valores son ceros, lo que sugiere días sin precipitación.

La media y mediana son bajas (4.4 y 0.3, respectivamente), lo que indica que la mayoría de las veces la precipitación es baja, pero hay valores altos (máximo de 91.8).

Existe una alta curtosis (22.7) y sesgo (4.07), lo que indica la presencia de eventos extremos de precipitación lo que quiere decir que la distribución de los datos de precipitación es muy desigual y tiene eventos extremos.

Temperatura: El 18% de los valores están ausentes, los valores van de 16 a 26.6, con una media y mediana de 21.4. la distribución de los datos es bastante simétrica, ya que tanto el sesgo como la curtosis están cercanos a 0.

Humedad: El 20% de los valores están ausentes, los valores varían de 50 a 100, con una media de 85.6 y una mediana de 86, la distribución es bastante simétrica.

Velocidad: El 44% de los valores están ausentes, el rango es bastante estrecho, de 0.88 a 3.17, con una media de 1.48 y una mediana de 1.44, la curtosis y el sesgo indican una ligera tendencia hacia valores más altos, pero no de manera extrema.

8.4. Observaciones:

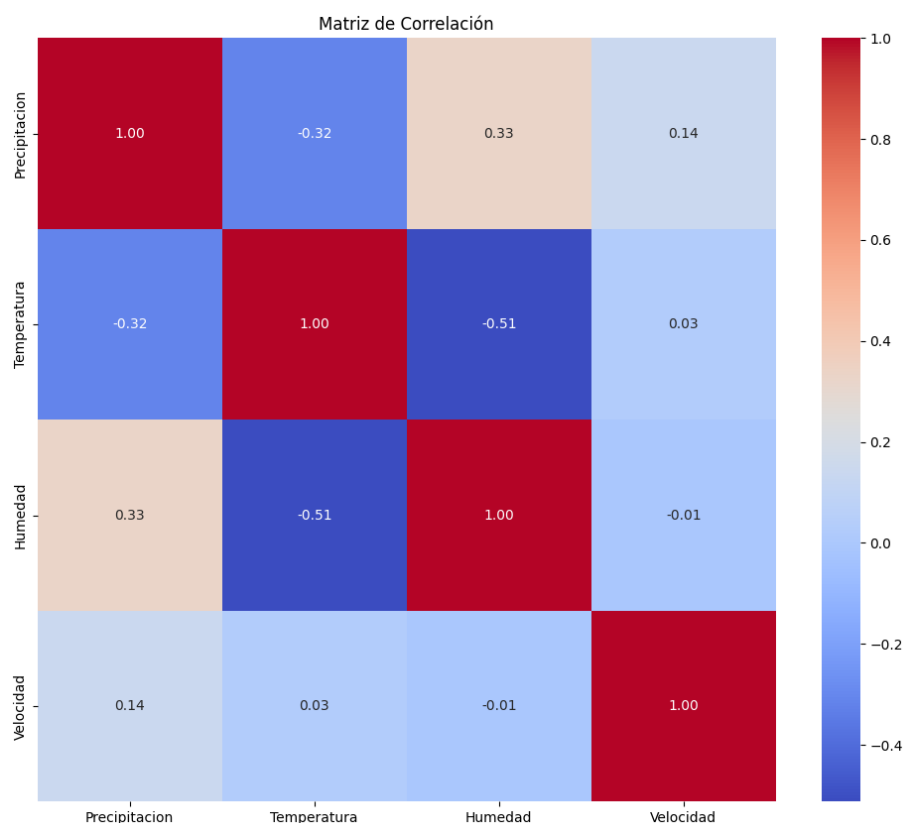
Valores faltantes: Existen valores faltantes significativos en varias variables, especialmente en Precipitación y Velocidad, esto puede afectar la calidad del análisis.

Distribución: Las variables como Nivel y Precipitación tienen sesgo y curtosis altos, lo que indica que los datos pueden tener comportamientos extremos o eventos infrecuentes.

Variables categóricas: Las variables categóricas, como Código Estación y Nombre Estación, no parecen aportar variabilidad relevante ya que tienen un solo valor por lo que se eliminan de la base.

8.5. Correlación de los datos

8.5.1. Matriz de correlación #1: Sin día y Mes y Sin Codificar.



Esta matriz de evaluación refleja la relación entre las variables **Precipitación**, **Temperatura**, **Humedad** y **Velocidad**. A continuación, se detallan las observaciones principales de esta evaluación:

La variable **Precipitación** tiene una presión positiva moderada con **Humedad** (0.33), lo que indica que, en días con más precipitación, la humedad tiende a ser mayor, presenta una compresión negativa moderada con **Temperatura** (-0.32), lo que indica que las temperaturas más altas podrían estar asociadas con menores niveles de precipitación. Así mismo, muestra una evaluación baja con **Velocidad** (0.14), lo que indica una relación poco significativa entre ambas variables.

La variable **Temperatura** muestra una correlación negativa moderada con **Humedad** (-0.51), indicando que a medida que la temperatura aumenta, la humedad tiende a disminuir, lo cual es consistente en ciertos climas donde el calor reduce la humedad ambiental, así mismo, tiene una relación baja con **Velocidad** (0.03), indicando que la temperatura y la velocidad no tienen una relación significativa en esta muestra de datos.

La variable **Humedad** presenta una correlación positiva moderada con **Precipitación** (0.33), lo cual es esperable, ya que una mayor precipitación suele estar acompañada de un incremento en la humedad y una correlación negativa moderada con **Temperatura** (-0.51), lo cual indica una relación inversa entre la humedad y la temperatura.

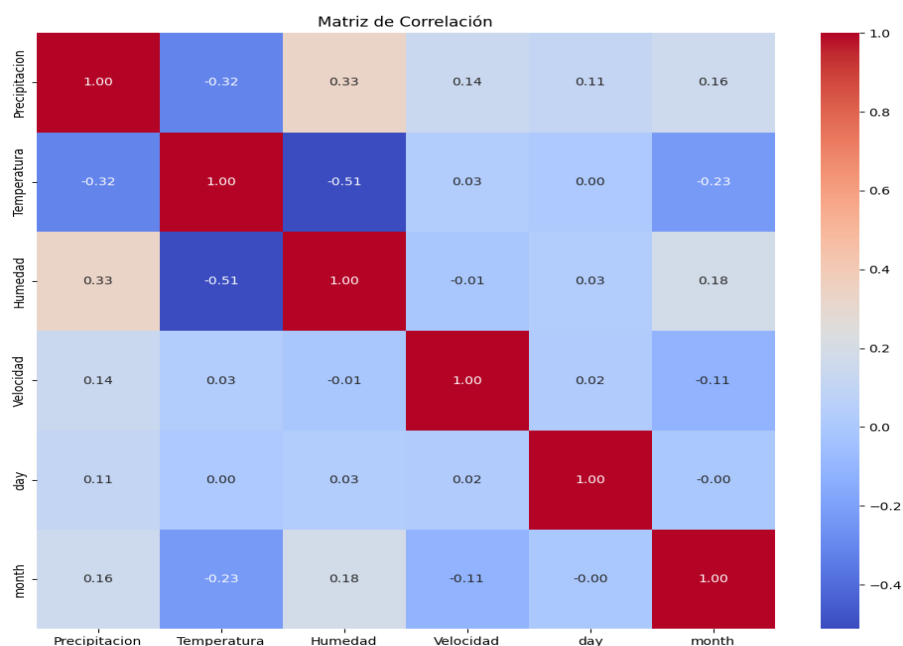
La variable **Velocidad** Tiene correlaciones muy bajas con **Precipitación** (0.14), **Temperatura** (0.03) y **Humedad** (-0.01), lo que indica que esta variable es casi independiente de las demás en esta muestra de datos.

La matriz indica una relación inversa moderada entre la **Temperatura** y la **Humedad**, lo que sugiere que el aumento en la temperatura podría reducir la humedad ambiental.

Una evaluación moderada positiva entre **Precipitación** y **Humedad**, lo cual es coherente con las condiciones atmosféricas de un clima húmedo asociado a precipitaciones.

Velocidad no muestra una valoración significativa con las otras variables, indicando que, en este caso, es una variable independiente en gran medida y que no aporta al modelo.

8.5.2. Matriz de correlación # 2 día y mes codificados con variable velocidad



La matriz de correlación muestra la relación entre diferentes variables meteorológicas tales como: **Precipitación, temperatura, humedad, velocidad, día, mes** que se describen a continuación:

La variable **Precipitación** tiene una presión moderada positiva con **Humedad** (0.33), lo que indica que, al aumentar la precipitación, la humedad también tiende a aumentar y una correlación negativa leve con **Temperatura** (-0.32), lo que sugiere que las temperaturas más altas podrían estar relacionadas con menos precipitaciones.

La variable **Temperatura** presenta una correlación negativa moderada con **Humedad** (-0.51), indicando que temperaturas más altas tienden a estar asociadas con una menor humedad en el ambiente.

Relación leve negativa con el **Mes** (-0.23), lo que podría implicar un ligero cambio en la temperatura en función de la época del año.

La variable **Humedad** representa una correlación positiva leve con **Precipitación** (0.33) y una correlación negativa moderada con **Temperatura** (-0.51), lo cual ya se mencionó antes y refuerza la tendencia de que mayor temperatura puede estar asociada con menor humedad.

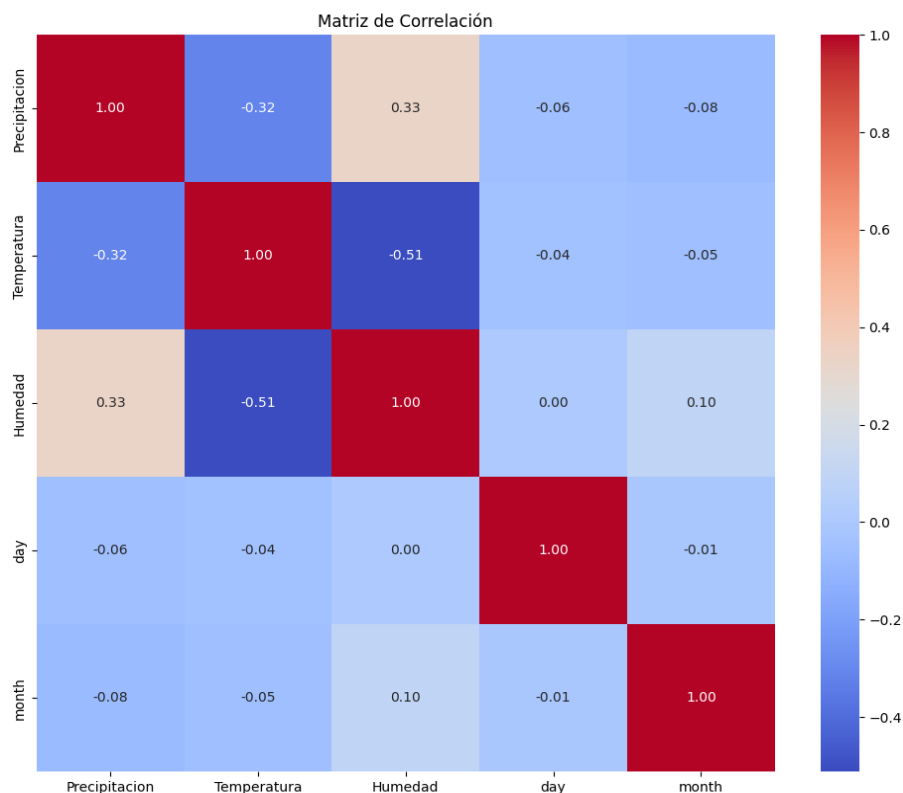
La variable **Velocidad** no muestra correlaciones significativas con ninguna de las demás variables, ya que todos sus valores de correlación son cercanos a cero. Esto indica que la velocidad no parece estar directamente relacionada con las otras variables en esta matriz, por lo que se decide eliminar esta variable del modelo.

La variable **Día y Mes** tienen una calificación de 1.0 consigo mismos (en la diagonal) y baja calificación con el resto de las variables meteorológicas, esto sugiere que los valores de día y mes no influyen fuertemente en las otras variables en esta muestra.

En general, la matriz muestra que la **Temperatura y Humedad** tienen la relación inversa más fuerte, lo cual es consistente con climas donde la humedad disminuye al aumentar la temperatura.

La **Precipitación** está asociada positivamente con la Humedad y negativamente con la **Temperatura**, sugiriendo que los días con precipitaciones podrían ser más frescos y húmedos.

8.5.3. Matriz de correlación #3: día y mes sin codificar y sin variable velocidad.



Esta matriz de correlación refleja la relación entre las variables **Precipitación**, **Temperatura**, **Humedad**, **Día** y **Mes**. A continuación, se detallan las observaciones principales de esta evaluación:

Existe una variación positiva moderada entre **Precipitación** y **Humedad** (0.33), lo cual indica que, en general, un aumento en la presión suele estar asociado con un incremento en la humedad. Esta relación es relevante, ya que la humedad del suelo y del ambiente podría ser un indicador de disponibilidad de agua en la cuenca.

Hay una compensación negativa relativamente fuerte entre **Temperatura** y **Humedad** (-0.51), lo que sugiere que a medida que la temperatura aumenta, la humedad tiende a disminuir. Esto es importante, ya que una menor humedad puede afectar la disponibilidad de agua en la cuenca, especialmente en períodos de alta temperatura.

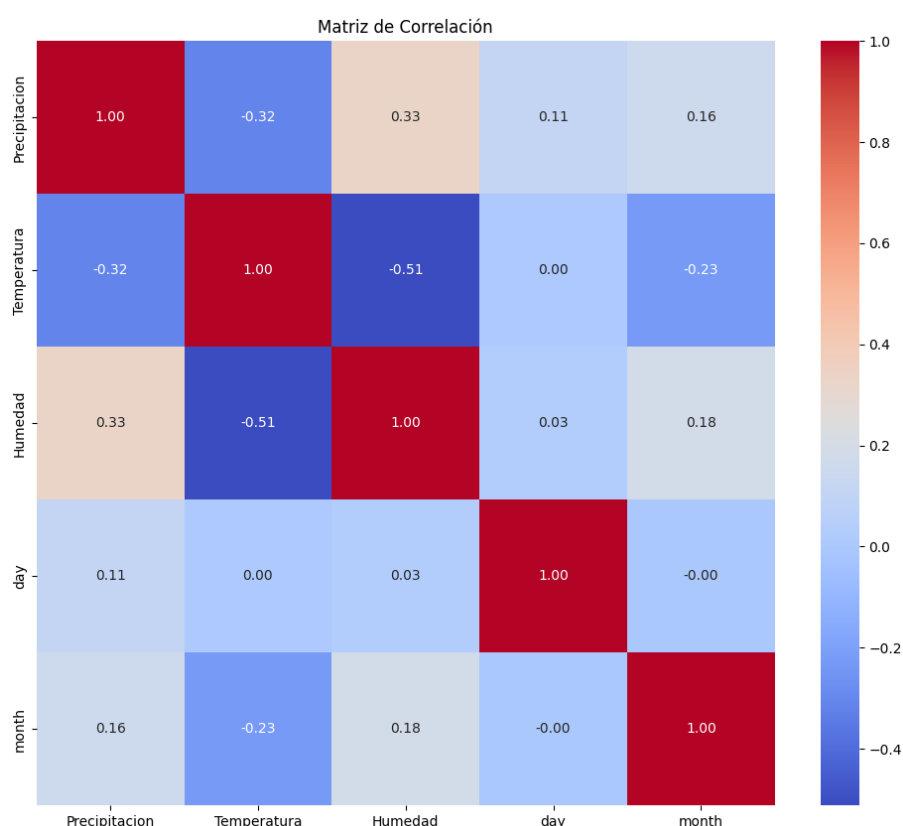
Existe una compensación negativa moderada entre **Precipitación** y **Temperatura** (-0.32), esta relación indica que las temperaturas tienden a ser más bajas cuando hay

alta precipitación, lo cual podría influir en los patrones de lluvia y, en consecuencia, en los recursos hídricos.

Día y Mes con las variables climáticas **Precipitación, Temperatura y Humedad** representan una correlación muy baja (entre -0.08 y 0.10), lo que sugiere que no existe una relación lineal clara entre las fechas y las condiciones climáticas en esta base de datos.

Esta matriz de estimación resalta algunas relaciones entre variables climáticas, especialmente entre temperatura y humedad, y entre precipitación y humedad. Estas correlaciones podrían usarse para identificar qué variables son más relevantes para el modelo predictivo y cómo podrían afectar la disponibilidad de agua en las cuencas del Municipio de Manizales.

8.5.4. Matriz de correlación #4 Día y Mes codificadas y sin variable velocidad.



La matriz de correlación muestra la relación entre las variables **Precipitación, Temperatura, Humedad, Día** (día codificado), y **Mes** (mes codificado), a continuación, se analizan las correlaciones relevantes entre estas variables:

Existe una precipitación positiva moderada entre **Precipitación y Humedad** (0.33), indicando que cuando aumenta la precipitación, la humedad tiende a subir, esta relación es importante para la predicción de recursos hídricos, ya que la humedad ambiental puede depender en parte de la cantidad de lluvia.

Hay una evaluación negativa significativa entre la **Temperatura** y la **humedad** (-0.51). Esto sugiere que a medida que la temperatura aumenta, la humedad disminuye. Este comportamiento es esperable, ya que las temperaturas altas suelen provocar una mayor evaporación, reduciendo la humedad relativa.

La compensación negativa entre **Precipitación y temperatura** (-0.32) indica que cuando hay más precipitación, la temperatura tiende a ser más baja, este efecto es común en áreas donde la lluvia refresca el ambiente.

La variable temporal **Día** tiene correlaciones muy bajas con las variables climáticas, lo cual indica que el día del mes no está linealmente relacionado con la precipitación, temperatura o humedad.

La variable temporal **Mes** muestra una medición baja con **temperatura** (-0.23) y **humedad** (0.18), sugiriendo que hay una leve tendencia de temperatura y humedad a variar a lo largo de los meses.

Este análisis de evaluación muestra que las relaciones entre **Precipitación, Temperatura y Humedad** pueden ser útiles para desarrollar un modelo de predicción de recursos hídricos, ya que estas variables tienen relaciones significativas entre sí. Las variables temporales **Día** y **Mes** parecen tener poca influencia directa en las condiciones climáticas diarias, lo que sugiere que pueden no ser determinantes por sí solas en el modelo, aunque podrían ayudar a capturar patrones estacionales al combinarse con otras variables.

8.6. Resultado de correlación.

Para la elección de la mejor correlación para el modelo predictivo de recursos hídricos en el municipio de Manizales, se analizó cuál de las matrices refleja mejor las relaciones útiles entre las variables climáticas (precipitación, temperatura, humedad y velocidad) y permite identificar patrones temporales (día y mes)

Matriz de Correlación #1 (sin día y mes y sin codificar):

Esta matriz es sencilla, enfocándose solo en las variables climáticas y muestra relaciones importantes, como la presión negativa entre temperatura y humedad (-0.51) y la presión positiva entre precipitación y humedad (0,33). Aunque identifica relaciones directas entre las variables, carece de un componente temporal (día/mes),

lo que limita su capacidad para capturar patrones estacionales, que pueden ser importantes en un modelo de predicción hídrica.

Matriz de Correlación #2 (día y meses codificados con variable de velocidad):

La codificación de día y mes permite capturar patrones estacionales. Sin embargo, la variable velocidad no tiene una comparación significativa con otras variables y podría no aportar valor. Por tanto, incluir la velocidad en el modelo podría agregar ruido en lugar de información relevante. Los patrones de evaluación entre las variables climáticas son consistentes, y la codificación de día y mes da lugar a posibles capturas estacionales, pero sin una relación fuerte.

Matriz de Correlación #3 (día y mes sin codificar y sin velocidad):

Excluye la velocidad, que no es muy correlativa, y usa día y mes sin codificar, lo que hace que los valores de correlación con el resto de variables climáticas se mantienen bajos (de -0,08 a 0,10). Aunque identifique las relaciones principales entre las variables climáticas, la falta de codificación temporal limita su capacidad para detectar patrones estacionales claros.

Matriz de Correlación #4 (día y mes codificados y sin velocidad):

Codifica día y mes para representar mejor los patrones estacionales. Excluye la variable velocidad, que no aporta valor significativo, y destaca correlaciones clave, como la variación moderada entre temperatura y humedad (-0,51) y entre precipitación y humedad (0,33). Esta configuración permite capturar patrones temporales sutiles (como la ligera variación estacional de temperatura y humedad).

Resultado:

La Matriz de Correlación #4 es la más adecuada porque, al codificar las variables de día y mes y omitir la velocidad, facilita la detección de patrones estacionales sin agregar ruido innecesario. Esta estructura es más eficiente, destacando las correlaciones climáticas importantes y permitiendo que el modelo capture variaciones a lo largo del tiempo, fundamentales para la predicción de recursos hídricos en un entorno influenciado por la estacionalidad como el del municipio de Manizales.

9. Resultados del Modelo

9.1. Regresión Lineal: Este modelo busca una relación directa entre las variables (como precipitación, temperatura y humedad) y el recurso hídrico. Utiliza una línea recta para hacer las predicciones, lo cual es útil si las variables tienen una relación lineal o proporcional con el recurso hídrico.

9.2. Regresión Polinómica: A diferencia de la regresión lineal, este modelo emplea una ecuación polinómica, lo que le permite ajustarse a relaciones más complejas y curvas en los datos. Esto es útil si la relación entre las variables y el recurso hídrico no es estrictamente lineal y muestra cambios más complejos.

9.3. Random Forest: Es un modelo avanzado que utiliza múltiples árboles de decisión para mejorar la precisión. Cada árbol se entrena en diferentes partes del conjunto de datos y sus resultados se combinan. Esto hace que el modelo sea robusto y capaz de manejar y mejorar la variabilidad y las interacciones complejas entre las variables, lo cual es ideal para datos hidrológicos.

9.4. Árbol de Decisión: Este modelo funciona dividiendo los datos en segmentos basados en condiciones específicas, como si fueran ramas de un árbol. Es fácil de interpretar, ya que muestra cómo se toma cada decisión en función de las variables, aunque puede ser menos preciso que otros métodos en problemas complejos.

9.5. Gradient Boosting: Este modelo es similar a Random Forest, pero construye los árboles de decisión de manera secuencial, corrigiendo los errores de los árboles anteriores. Esto le permite mejorar constantemente su precisión y manejar mejor la complejidad de las relaciones entre variables, haciendo ideal para predicciones hidrológicas precisas.

9.6. Tabla de Resultados

TABLA DE RESULTADOS						
VARIABLES	PARTICIÓN	REGRESIÓN LINEAL	REGRESIÓN POLINÓMICA	RANDOM FOREST	ÁRBOL DE DECISIÓN	GRADIENT BOOSTING
Precipitación/temperatura/humedad/ velocidad	0.30	0.28	0.25 Grado 2	0.26	-0.48	0.22
Precipitación/ temperatura/humedad/día/ mes (numérico)	0.30	0.25	0.296 Grado 2	0.49	0.09	0.36
Precipitación/ temperatura/humedad/día/ mes (numérico)	0.20	0.29	0.35	0.52	-0.25	0.46
Precipitación/ temperatura/humedad/día/mes/target encoding	0.30	0.48	0.49 Grado 2	0.47	0.006X	0.44
Precipitación/ temperatura/humedad/día/mes/target encoding	0.20	0.48	0.47 Grado 2	0.48	0.06X	0.5
Precipitación/ temperatura/humedad/día/ mes/ target escalado + velocidad	0.30	0.47	0.02	0.47	-0.10	0.44

Precipitación/ temperatura/ humedad/día/Mes/ Leave encoding sin velocidad	0.30	0.46	0.47	0.76	0.64	0.69
---	------	------	------	------	------	------

Los modelos analizados en la tabla han mostrado variaciones significativas en su desempeño al usar diferentes variables y técnicas de procesamiento de datos. Los resultados indican que:

Random Forest y **Gradient Boosting** muestran consistentemente buenos resultados en casi todas las particiones de datos, especialmente cuando se incorporan transformaciones avanzadas como el left-one-out encoding (con valores de 0.76 y 0.69, respectivamente) y target encoding en algunas particiones.

Regresión Polinómica de Grado 2 también ha logrado un buen desempeño en varias combinaciones de variables, especialmente en configuraciones donde se utilizan técnicas de codificación. Sin embargo, la efectividad de este modelo disminuye en ciertos ajustes, como cuando se incluye la velocidad.

Regresión Lineal y **Árbol de Decisión** presentan un desempeño más limitado, con valores bajos y negativos en algunas particiones, sugiriendo que estos métodos podrían no capturar adecuadamente la complejidad de la interacción entre variables en el contexto hidrológico de las cuencas.

En conclusión, se observa que **Random Forest** es el modelo que mejor se adapta a la predicción de los recursos hídricos bajo este conjunto de variables y transformaciones. Por tanto, se opta por este modelo para optimizar la precisión del modelo de predicción, con un enfoque adicional para mejorar la selección y el procesamiento de las variables utilizadas.

9.7. Impactos y líneas futuras: El proyecto de predicción del Recurso Hídrico en las Cuencas del Municipio de Manizales genera impactos significativos en la gestión sostenible del agua y en la adaptación al cambio climático. Entre los principales impactos se encuentra la mejora en la capacidad de las autoridades locales para anticipar períodos de escasez o abundancia de agua, lo cual facilita una planificación hídrica efectiva y permite a la comunidad beneficiarse de un abastecimiento más seguro. Además, este proyecto aporta datos fundamentales para proyectos de energía renovable, especialmente hidroeléctricos, al proporcionar pronósticos precisos que informan sobre la viabilidad de estos desarrollos en términos de disponibilidad de agua.

En el ámbito de la investigación, el proyecto representa un avance para la ciencia de datos y la inteligencia artificial aplicada en Colombia, destacando el uso de metodologías avanzadas para resolver problemas ambientales críticos. A nivel nacional, establece un modelo que puede ser replicado en otras regiones del país, aumentando la resiliencia ante fenómenos climáticos extremos y contribuyendo a una gestión sostenible de los recursos hídricos.

El proyecto abre múltiples líneas de investigación y desarrollo, una de ellas consiste en integrar nuevas variables ambientales, como la cobertura del suelo y el uso agrícola, para refinar la precisión del modelo predictivo. También se puede explorar el uso de redes neuronales profundas que permitan mejorar el desempeño del modelo en condiciones de mayor variabilidad climática. Finalmente, el establecimiento de una plataforma en línea de acceso público podría transformar los resultados en una herramienta consultable por diversos sectores, promoviendo una gestión comunitaria y colaborativa del recurso hídrico.

10. Anexos

- 10.1. Procesamiento_base.ipynb
- 10.2. Manual de usuario líneas de Código
- 10.3. Análisis EDA
- 10.4. Matrices de correlación

11. Bibliografía

- EAM
Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM). (Año).
Título del informe o base de datos utilizados. IDEAM. URL de acceso (si aplica).
- Datos Públicos
Repositorios de datos públicos.
- Literatura científica, artículos
- Plan de Contingencia- Segunda Temporada de Lluvias 2024 Alcaldía de Manizales, Unidad de Gestión del Riesgo