Para el desarrollo del proyecto de predicción de recursos hídricos en las cuencas del Municipio de Manizales, seguimos la metodología CRISP-DM, la cual nos permitió estructurar y ejecutar el proyecto de manera eficiente. El proceso se llevó a cabo de la siguiente manera:

**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.

**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.

**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.

**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.

**5. Evaluación**

* Evaluamos el rendimiento del modelo utilizando métricas como el error cuadrático medio (MSE) y el coeficiente de determinación (R²), 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.

**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.

**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.

### **Comprensión del Problema**

* **Objetivo del Proyecto**: Definir el objetivo específico del modelo, que es predecir el nivel de las cuencas hídricas (e.g., nivel en metros cúbicos).
* **Contexto del Proyecto**: Explicar por qué es relevante predecir los niveles de cuencas (e.g., apoyo en la gestión de recursos hídricos, prevención de inundaciones).
* **Requerimientos y Limitaciones**: Identificar limitaciones, como tiempo, acceso a datos y recursos computacionales.
* **Métricas de Éxito**: Definir métricas clave de evaluación (e.g., error cuadrático medio) que indiquen el rendimiento del modelo.

### **2. Comprensión de los Datos**

* **Fuentes de Datos**: Explicar las fuentes disponibles (e.g., datos históricos de precipitaciones y nivel de agua).
* **Descripción de Variables**: Detallar cada variable (e.g., precipitación, temperatura) y su posible impacto en el nivel de la cuenca.
* **Análisis Exploratorio de Datos (EDA)**: Realizar una exploración inicial para observar patrones y relaciones.
* **Evaluación de Calidad de Datos**: Documentar problemas de calidad como valores nulos, errores o datos extremos.

### **3. Preparación de los Datos**

* **Limpieza de Datos**: Manejar valores nulos, duplicados y atípicos que podrían afectar el modelo.
* **Selección de Variables**: Determinar las variables más relevantes para la predicción.
* **Transformación de Variables**: Escalar y normalizar los datos; crear variables derivadas si se necesitan.
* **División de los Datos**: Separar los datos en conjuntos de entrenamiento, validación y prueba para evaluar el modelo.

### **4. Modelado**

* **Selección de Algoritmos**: Probar con diferentes algoritmos (e.g., regresión lineal, árboles de decisión, redes neuronales) para encontrar el que mejor se adapte.
* **Ajuste de Modelos**: Entrenar cada modelo con un ajuste inicial y probar variaciones en sus parámetros.
* **Evaluación de Modelos**: Evaluar el rendimiento de cada modelo con métricas como RMSE (raíz del error cuadrático medio).
* **Optimización**: Ajuste fino y selección del mejor modelo para mejorar precisión.

### **5. Evaluación**

* **Validación Final**: Validar el rendimiento del modelo seleccionado con el conjunto de prueba.
* **Revisión de Métricas**: Asegurarse de que las métricas cumplen con los objetivos de precisión definidos.
* **Análisis de Errores**: Identificar posibles causas de error y limitaciones en el modelo.
* **Interpretación de Resultados**: Evaluar si los resultados son útiles para el contexto del proyecto.

### **6. Despliegue Simulado**

* **Plan de Implementación**: Explicar cómo se podría desplegar el modelo en un entorno real.
* **Prueba en Entorno de Práctica**: Probar el modelo en un entorno de simulación.
* **Monitoreo Simulado**: Describir cómo se supervisaría el rendimiento del modelo una vez en uso.
* **Documentación**: Elaborar un informe técnico con todos los pasos y resultados.

### **7. Conclusiones y Recomendaciones**

* **Resumen del Proyecto**: Recapitular los hallazgos y el rendimiento del modelo.
* **Retos y Soluciones**: Documentar dificultades encontradas y soluciones aplicadas.
* **Posibles Mejoras**: Sugerir mejoras para futuros ajustes del modelo o para mejorar la calidad de los datos.