

Predicciones de nivel de presas en el estado de Sonora

Autores:

Luis Hiram Hernández Gutiérrez (a219210858@unison.mx)

Francisco Yáñez Bustamante (a220211785@unison.mx)

Resumen:

Este proyecto se centró en anticipar los niveles de agua en diversas presas distribuidas a lo largo del estado de Sonora, México. Aprovechando técnicas de inteligencia artificial, logramos desarrollar un modelo capaz de prever los niveles de agua en una presa en una fecha específica. La implementación se llevó a cabo en el lenguaje de programación Python, haciendo uso de bibliotecas fundamentales como Pandas, Numpy, Matplotlib y Scikit-Learn. Además, incorporamos el modelo XGBoost para realizar predicciones basadas en diversas características. Esta combinación de herramientas nos permitió confeccionar un modelo que proporciona respuestas precisas a la pregunta de cómo evolucionarán los niveles de agua en las presas, contribuyendo así a una gestión más eficiente de los recursos hídricos en la región.

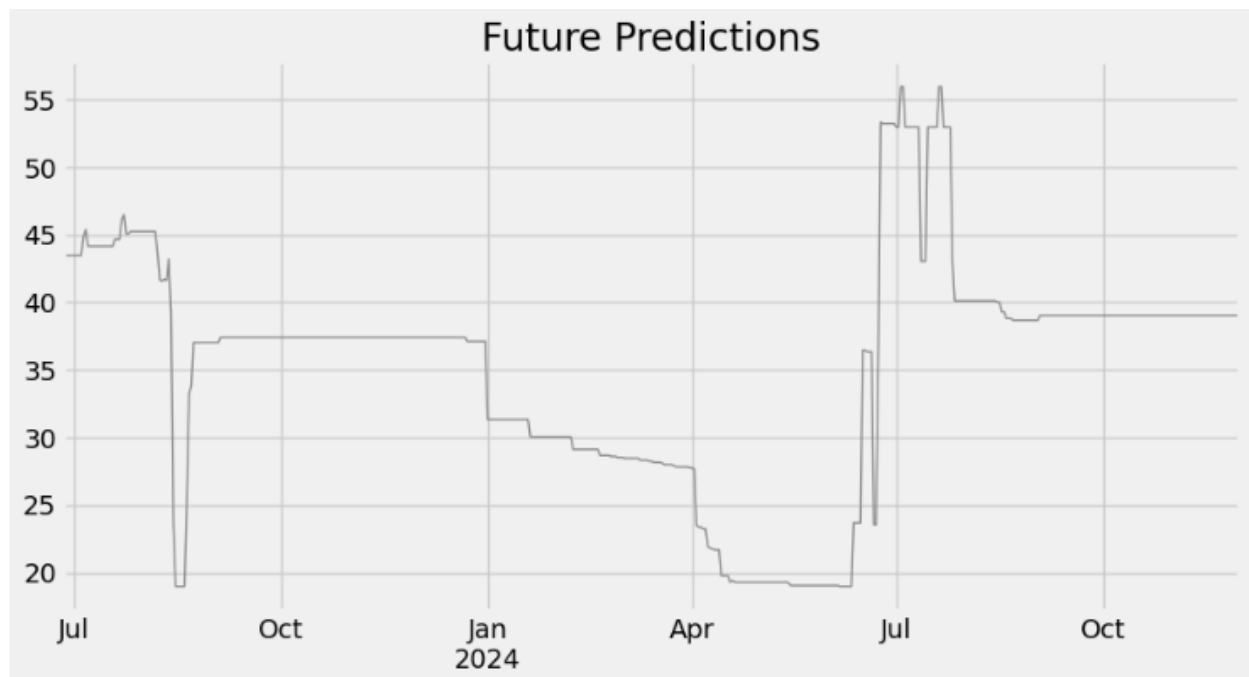


Ilustración 1: Ejemplo de la salida del modelo. Predicción para presa con clave "AGZCH" para el año 2024



Desarrollo:

Para llevar a cabo este proyecto, inicialmente nos basamos en los conjuntos de datos proporcionados, los cuales contenían información sobre los niveles de 10 presas distintas en el estado desde enero de 2020 hasta junio de 2023, suministrados por el gobierno. Reconociendo la importancia de la cantidad y calidad de los datos en el rendimiento del modelo, decidimos complementar esta información mediante la búsqueda de datos adicionales en páginas especializadas.

Tras analizar varias fuentes de datos y descartar algunas, identificamos una base de datos en VisualCrossing, una plataforma dedicada a mediciones climatológicas. En este sitio, obtuvimos información climática de Hermosillo durante los últimos 13 años, incluyendo variables como temperatura, humedad, registros de lluvia, nubosidad, entre otras. Estas características adicionales enriquecieron nuestro conjunto de datos y proporcionaron una perspectiva más completa para el análisis.

Inicialmente, llevamos a cabo un análisis exploratorio de los datos para asegurarnos de su utilidad, eliminando características no relevantes, como "nieve", y abordando datos nulos mediante la detección y llenado de los mismos. Una vez completada esta fase de "limpieza" de datos, procedimos a la integración de ambos conjuntos, creando así un nuevo conjunto de datos consolidado que sería utilizado para entrenar nuestro modelo predictivo. Este enfoque combinado nos permitió aprovechar la información de múltiples fuentes y mejorar la robustez de nuestro modelo.

	clave	fecha	almacenamiento	temp	humidity	precip	preciptype	uvindex	conditions
0	AGZCH	2020-01-01	79.100	14.7	65.0	0.0	0	6	Clear
1	AOBSO	2020-01-01	1683.920	14.7	65.0	0.0	0	6	Clear
2	ARCSO	2020-01-01	501.460	14.7	65.0	0.0	0	6	Clear
3	ARLSO	2020-01-01	18.420	14.7	65.0	0.0	0	6	Clear
4	CHTSO	2020-01-01	13.170	14.7	65.0	0.0	0	6	Clear
...
11528	CHTSO	2023-06-26	2.235	35.7	21.2	0.0	0	8	Partially cloudy
11529	IRASO	2023-06-26	0.295	35.7	21.2	0.0	0	8	Partially cloudy
11530	LCDSO	2023-06-26	547.737	35.7	21.2	0.0	0	8	Partially cloudy
11531	PECOS	2023-06-26	959.204	35.7	21.2	0.0	0	8	Partially cloudy
11532	PMOSO	2023-06-26	25.044	35.7	21.2	0.0	0	8	Partially cloudy

11533 rows × 9 columns

Ilustración 2: Datos ya procesados. Podemos notar que son un total de 11533 filas y 7 características sin contar clave y fecha



Continuando con el proyecto, nos embarcamos en la creación de nuestro modelo predictivo. En esta instancia, optamos por emplear XGBoost, una elección respaldada por su capacidad para manejar conjuntos de datos complejos y su eficacia demostrada en la predicción de series temporales. Además, habíamos previamente trabajado con este modelo en proyectos anteriores, lo que contribuyó a nuestra confianza en su desempeño.

Para garantizar una evaluación robusta de nuestro modelo, procedimos a dividir nuestros datos en conjuntos de prueba y validación. Implementamos un 3-fold cross validation, lo cual nos permitió obtener estimaciones más fiables y precisas.

Después de entrenar y validar nuestro modelo, utilizamos RMSE para cuantificar la calidad de las predicciones en el conjunto de datos de validación. Esta métrica nos proporcionó una medida de la precisión del modelo en términos de la diferencia entre las predicciones y los valores reales de los niveles de presas.

Con la validación completa, nos dirigimos hacia la fase de predicción del modelo para el año 2024. Dado que nuestros datos históricos llegaban hasta junio de 2023, empezamos las predicciones desde esa fecha hasta diciembre de 2024.

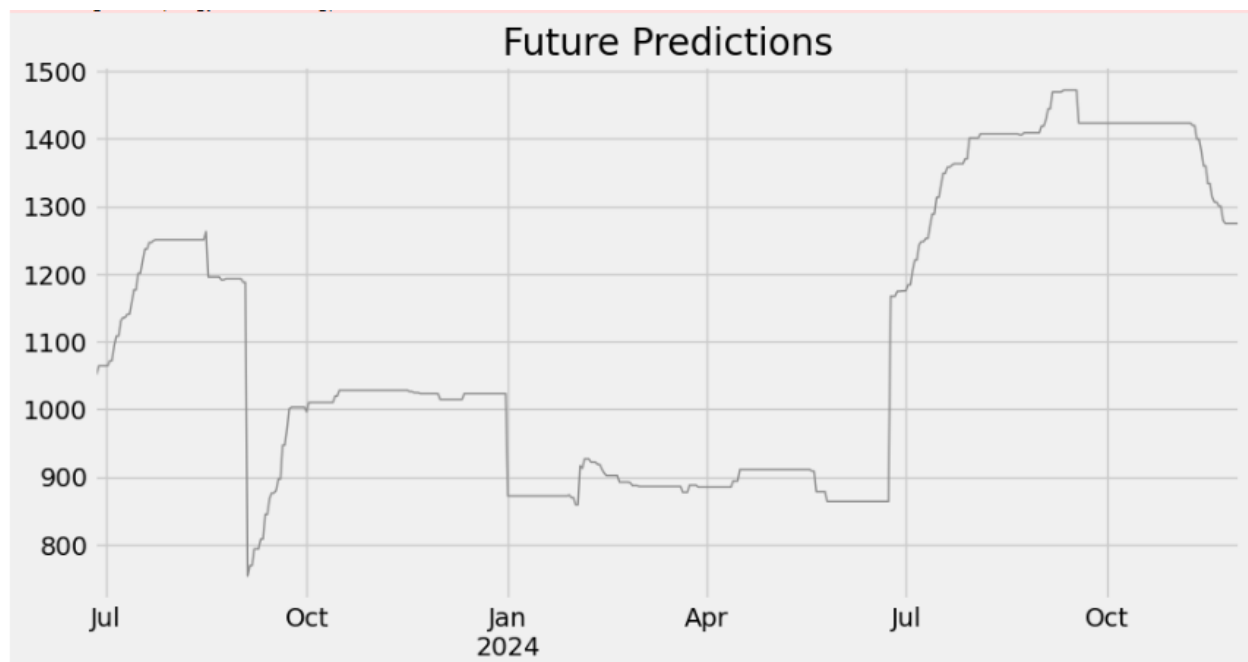


Ilustración 3: Predicciones para presa con clave "AOBSO". Podemos observar que efectivamente predice la subida del nivel de las presas en temporada de lluvias.