

PREVISIÓN DE GENERACIÓN HIDRÁULICA EN EL CONTEXTO DE TRANSICIÓN ENERGÉTICA Y VOLATILIDAD DEL MERCADO

Héctor González - Pablo Caldevilla - Jonathan Ponce|Tutor académico: Stella Maris Salvatierra Galiano|Tutor de empresa: Marta Enesco Garrido|Cotutor de empresa: Roberto Flores Herrera

Universidad de Navarra - Máster en Big Data Science

Resumen

Este proyecto evalúa técnicas de *Machine Learning* (XGBoost, LightGBM), modelado tradicional (SARIMAX) y enfoques alternativos (Prophet) para pronosticar el caudal diario del río Duero en Zamora, España, con un horizonte de predicción de 7 días; tomando como referencia a la central hidroeléctrica de Pereruela y San Román. Se emplearon datos del río Duero obtenidos mediante *web scraping* y datos meteorológicos a través de API's *open-source*.

LightGBM mostró el mejor desempeño en la predicción de picos, mientras que XGBoost y SARIMAX fueron consistentemente estables en todas las métricas de error. Prophet, en cambio, obtuvo los peores resultados.

Las variables más influyentes en el comportamiento del caudal fueron la lluvia, la humedad del suelo, la cobertura de nubes, las ráfagas de viento y la temperatura.

Obejtivos

Desarrollar un sistema integral de pronóstico diario para el caudal del río Duero en Zamora, en un horizonte semanal (7 días), utilizando técnicas avanzadas de modelado tradicional y aprendizaje automático, generando una comparativa de modelos y enfoques estadísticos.

- 1) Diseñar un flujo de extracción, limpieza y análisis de series temporales para datos históricos del caudal del Duero en Zamora y variables meteorológicas relevantes.
- 2) Implementar modelos de series de tiempo, SARIMAX, Prophet, XGBoost y LightGBM para el pronóstico del caudal del Duero en Zamora. Evaluar y comparar su desempeño en términos de error de pronóstico y captura de picos o subidas repentinas.
- 3) Analizar e interpretar los factores internos y externos que afectan el comportamiento del caudal del río Duero en Zamora.

Obtención de los datos

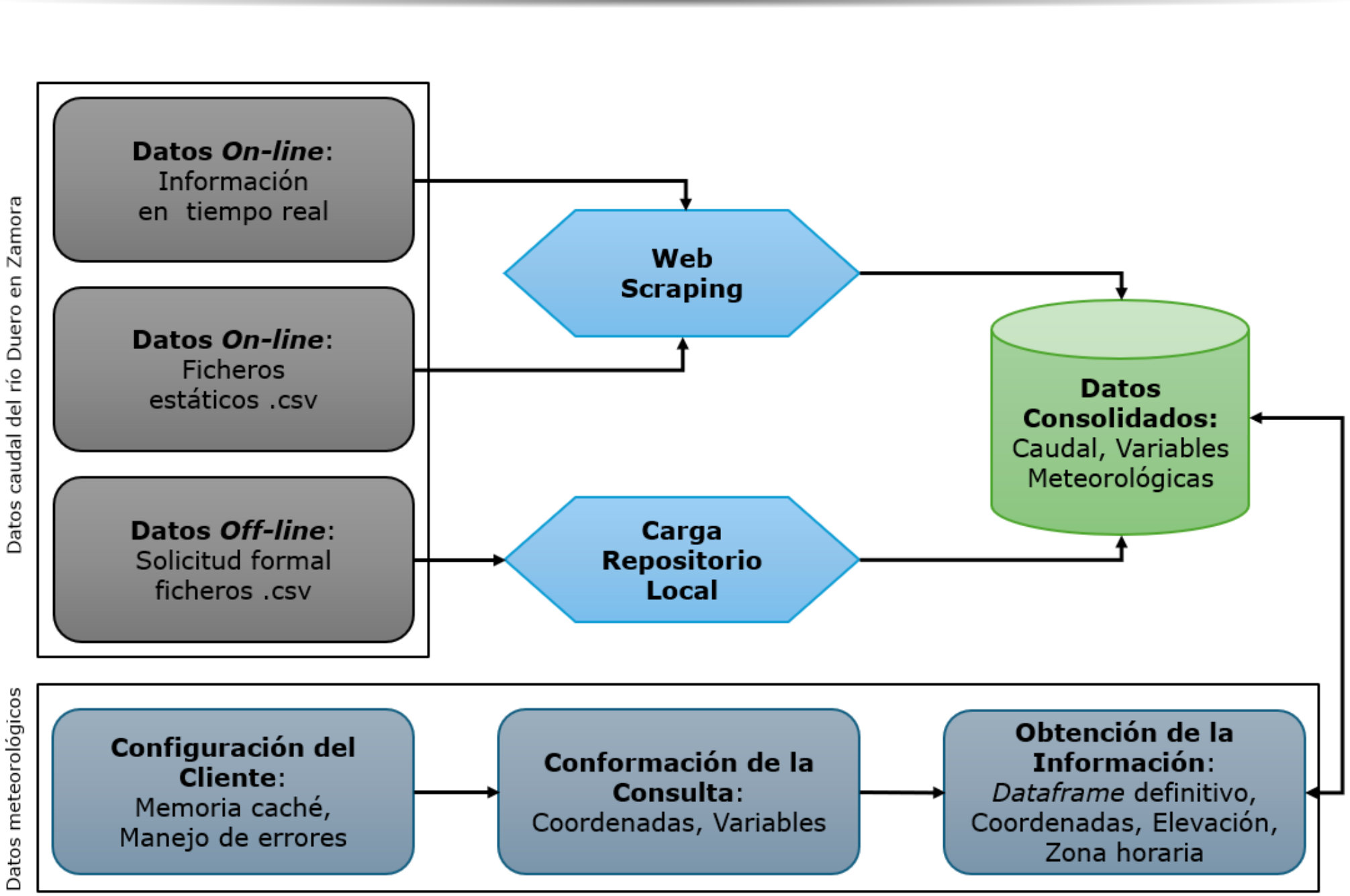


Figure 1:Diagrama del flujo de extracción de la información.

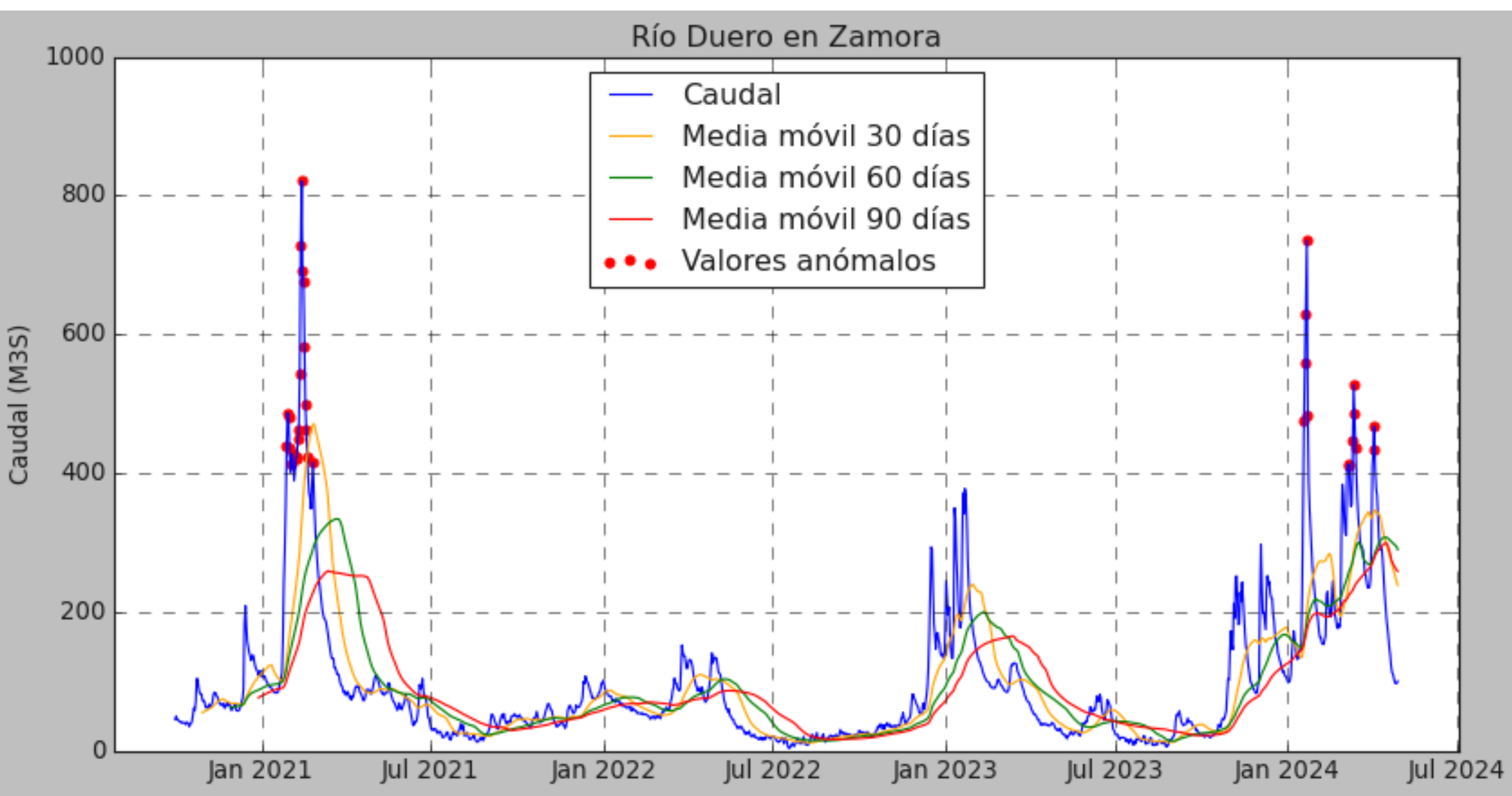


Figure 2:Comportamiento del caudal del río Duero en Zamora.

Entrenamiento de los modelos

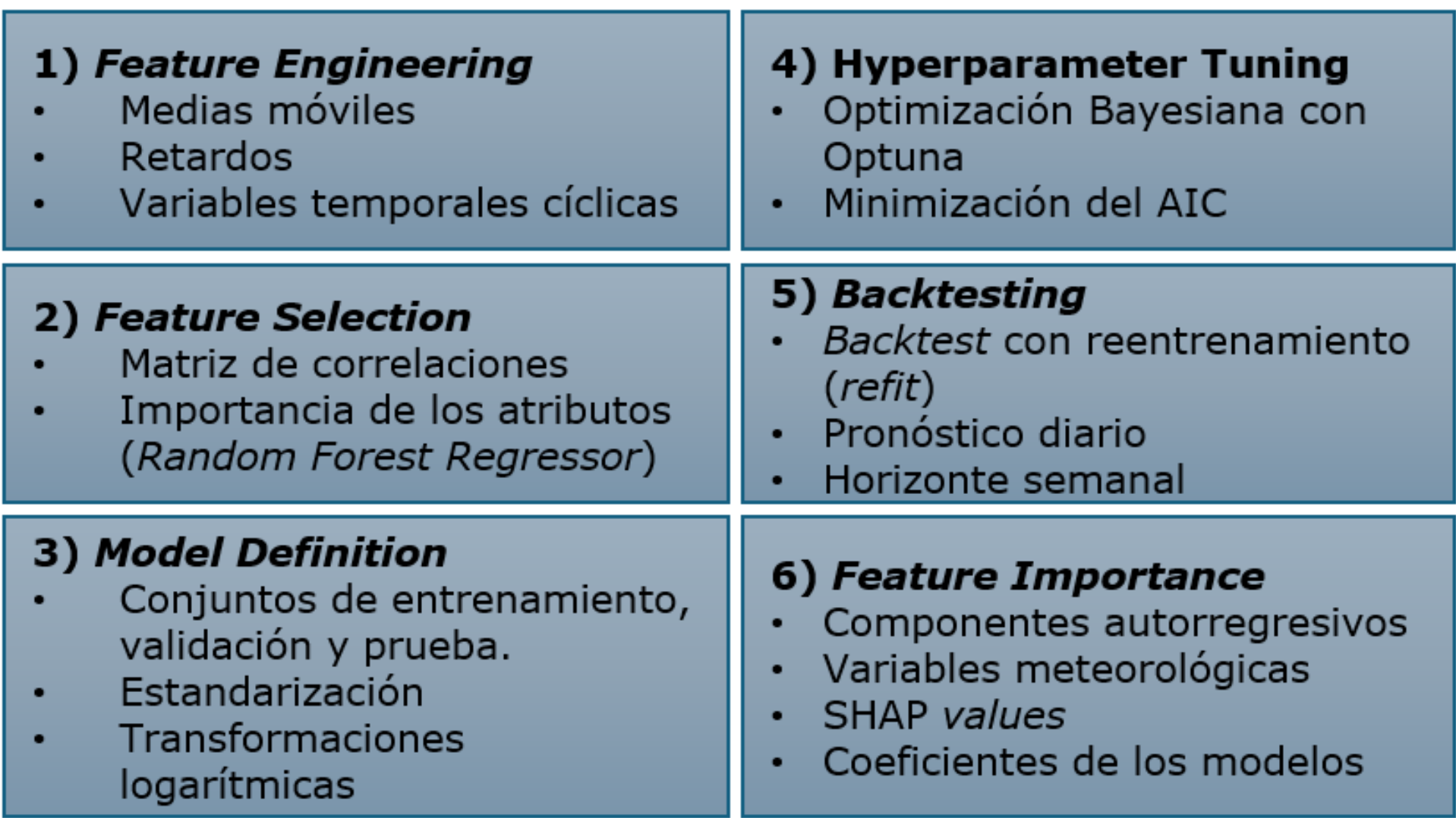


Figure 3:Flujo de entrenamiento de los modelos.

El *backtesting* con *refit* permitió poner a prueba en un entorno simulado a los cuatro algoritmos y encontrar el modelo ganador para cada métrica de error, en especial aquella que evalúa la precisión en la predicción de picos.

Resultados

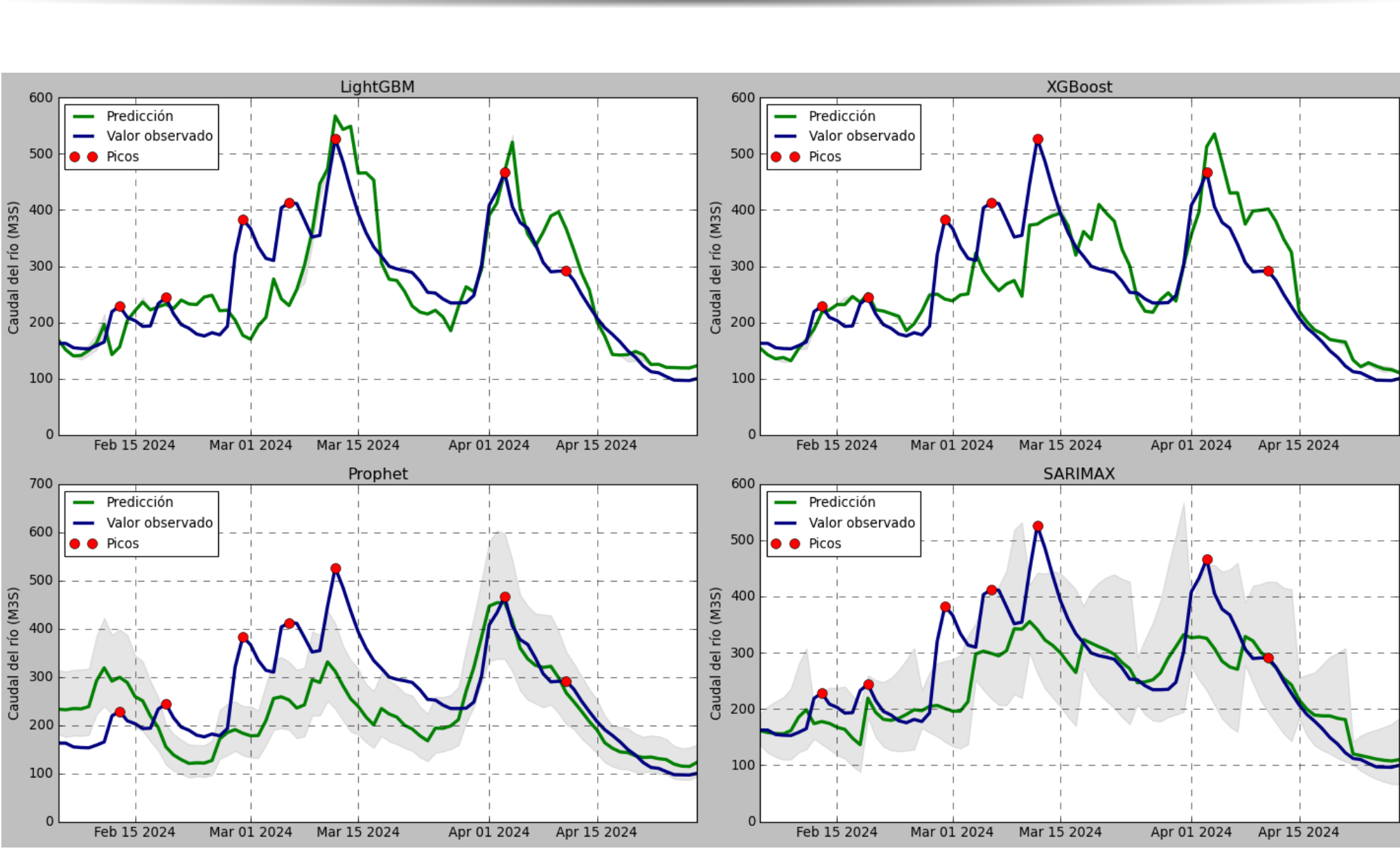


Figure 4:Backtesting en los modelos.

Metric	Baseline	LightGBM	XGBoost	Prophet	SARIMAX
MAE Peaks	108.64	84.48	87.16	110.40	100.15
MAPE	0.21	0.18	0.18	0.26	0.15
MSE	5155.51	4511.77	4332.46	7721.39	4350.58
RMSE	71.80	67.17	65.82	87.87	65.96

Table 1:Comparación del desempeño de los modelos.

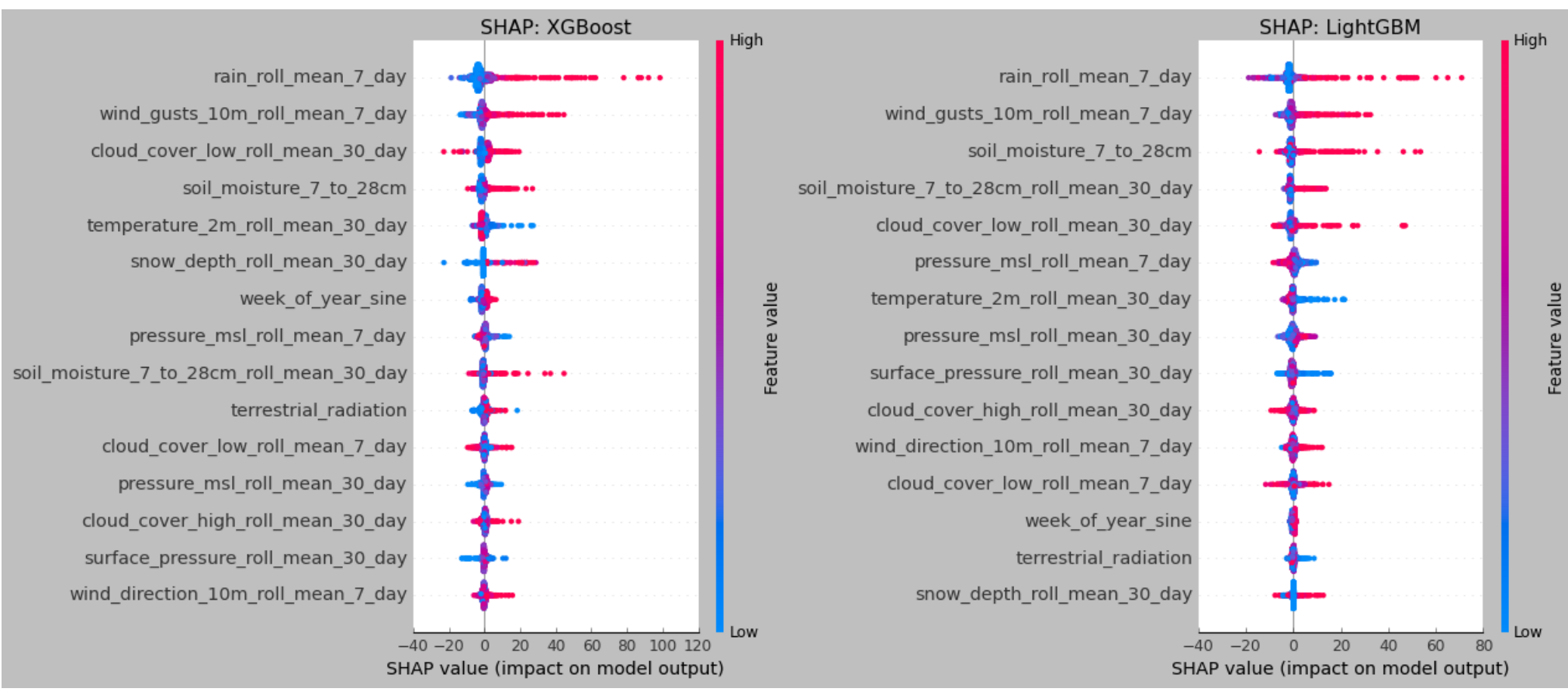


Figure 5:SHAP values modelos de ensemble.

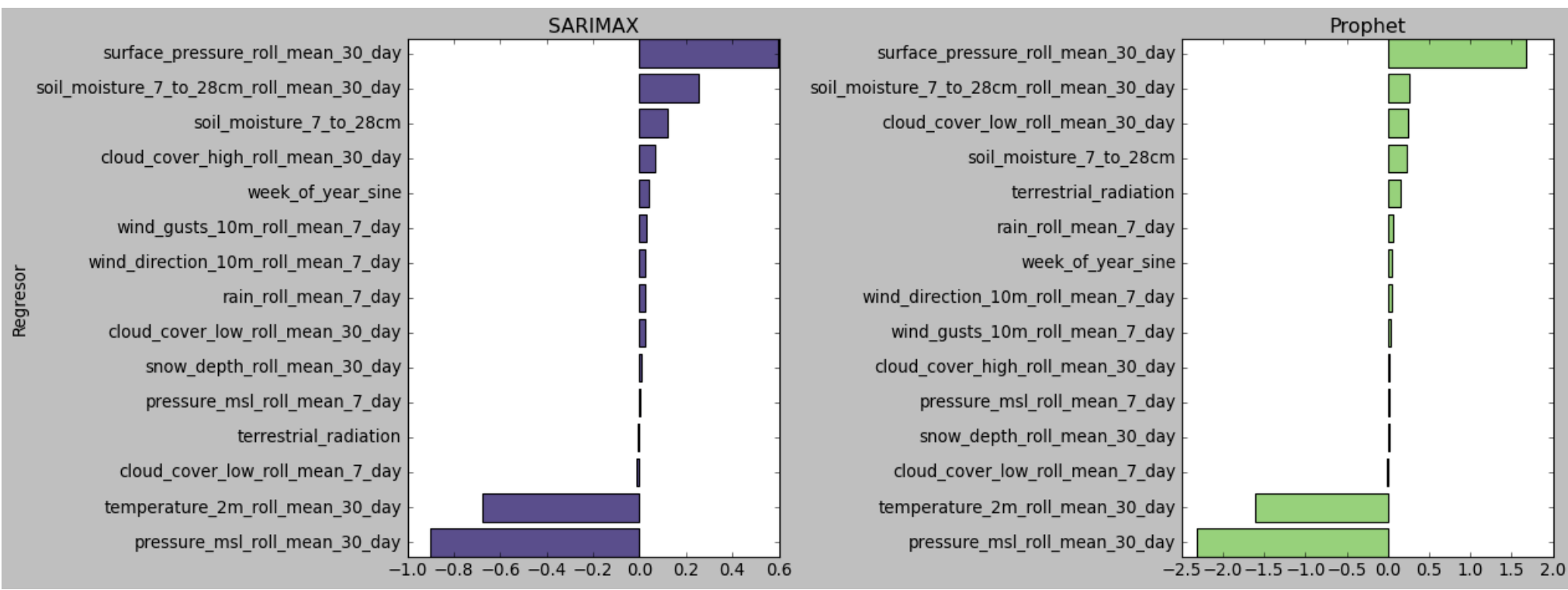


Figure 6:Coeficientes modelos SARIMAX y Prophet.

Para XGBoost y LightGBM, los retardos recientes (lag 1, lag 2 y lag 3) son cruciales para predecir el caudal, con lag 2 y lag 1 teniendo el mayor impacto (correlación positiva). En SARIMAX, el primer retardo (ar.L1) es significativo y su correlación con el caudal es positiva.

Conclusiones

Luego de finalizar el proyecto, los resultados permitieron llegar a conclusiones de carácter técnico, metodológico y de negocio:

- 1) La ingeniería de características y la selección de regresores han optimizado significativamente el conjunto de datos base. La inclusión de medias móviles de variables meteorológicas y la codificación cíclica han mejorado la precisión y la interpretabilidad de los modelos predictivos. El 85% de las variables seleccionadas para la estimación de los modelos correspondieron a modificaciones de los atributos originales.
- 2) En la evaluación de algoritmos, los modelos de ensemble se destacaron en la identificación de picos y anomalías del caudal. Aunque el modelo SARIMAX tuvo un nivel de error bajo en todas las métricas planteadas, no resaltó en la predicción de eventos atípicos. El modelo Prophet no logró superar el desempeño del modelo *base line*.
- 3) La capacidad de los modelos avanzados para pronosticar con precisión el caudal de los ríos es crucial para las compañías hidroeléctricas. Esto les permite anticipar la producción de energía y ajustar sus estrategias comerciales para maximizar ingresos. Al contar con información precisa y anticipada, las compañías pueden establecer precios de manera más estratégica, optimizando sus recursos y asegurando una oferta competitiva en el mercado de energías renovables.

Bibliografía

- <https://link.springer.com/article/10.1007/s12665-024-11481-w> Modeling and forecasting rainfall patterns in India: a time series analysis with XGBoost algorithm
- <https://www.tandfonline.com/doi/full/10.1080/19942060.2021.2019128> Water level prediction using various machine learning algorithms: a case study of Durian Tunggal river, Malaysia