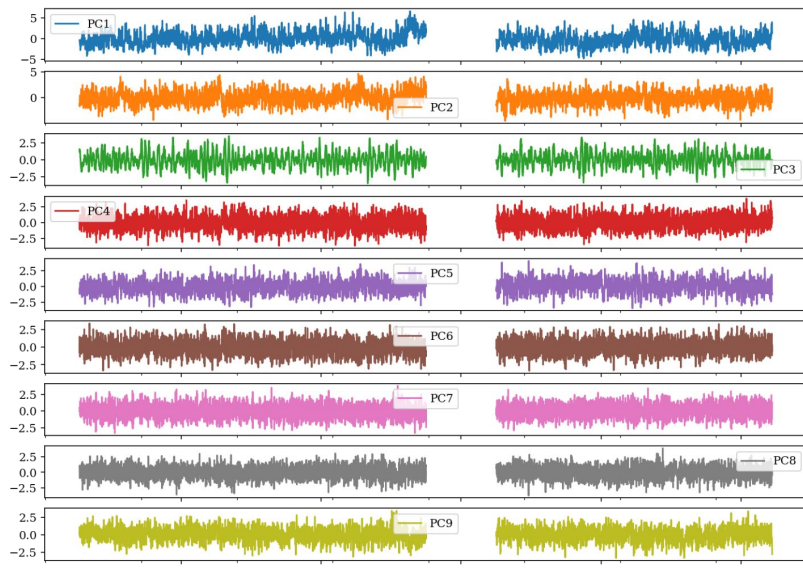


Predicción de caudales medios diarios usando modelos de inteligencia artificial y aprendizaje automático

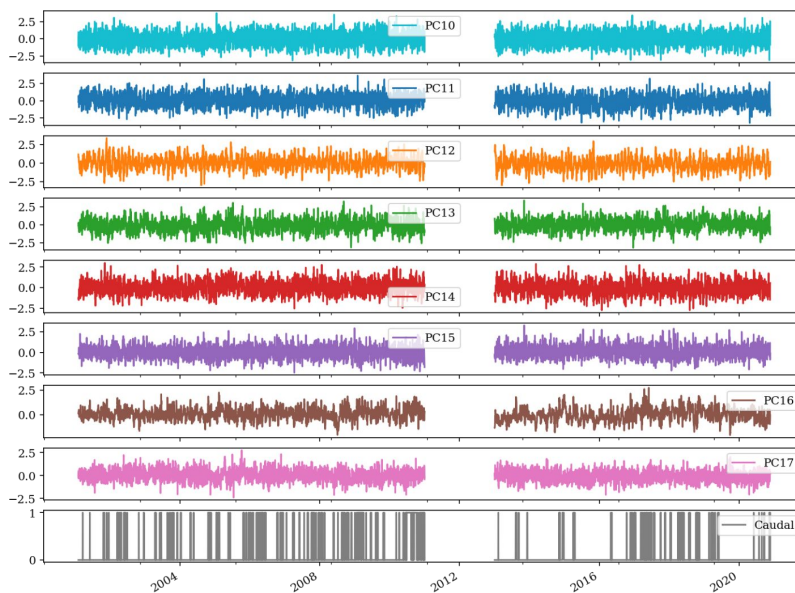
Clasificación de eventos extremos

Juan Esteban Taborda Soto

Series referentes a eventos máximos de caudal

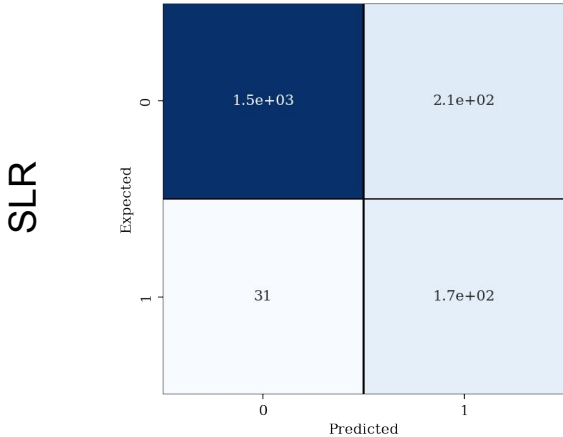
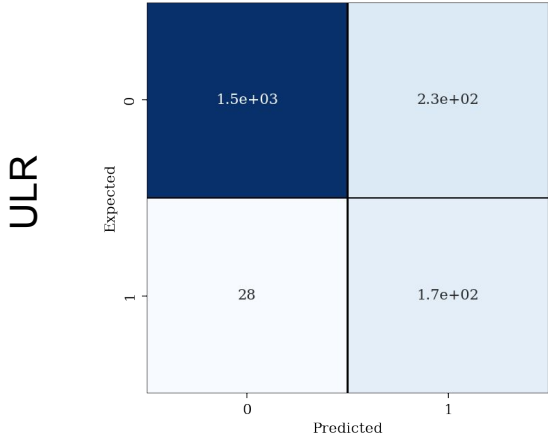
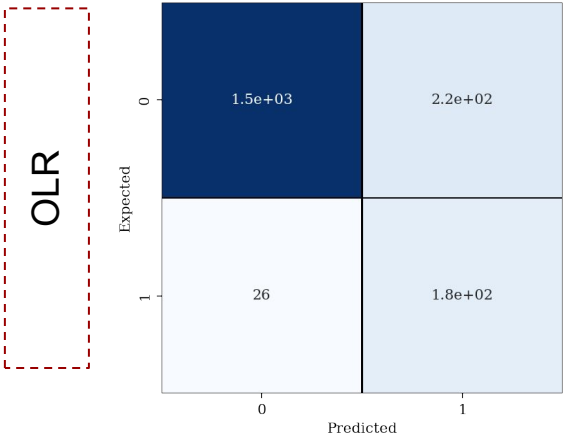
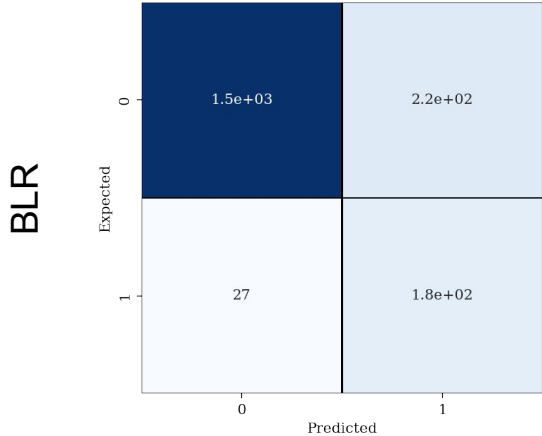
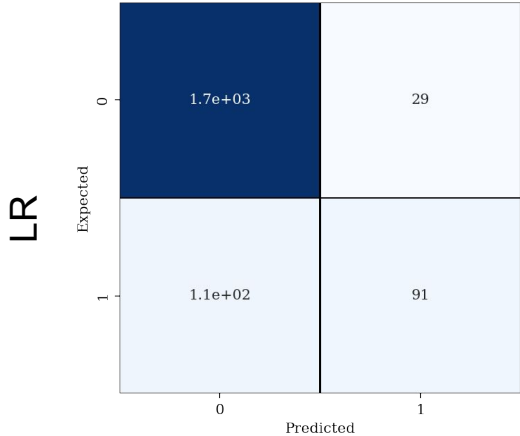


Existe desbalance entre los valores máximos ($\text{obs} \geq p_{90}$) y otros

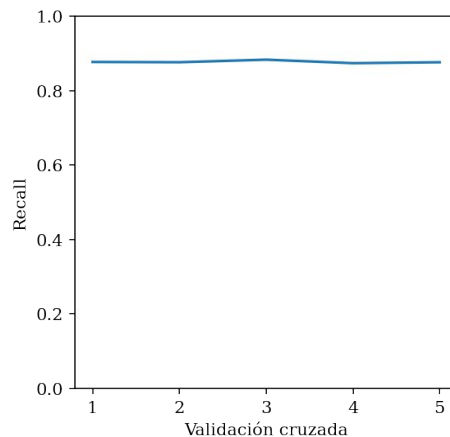
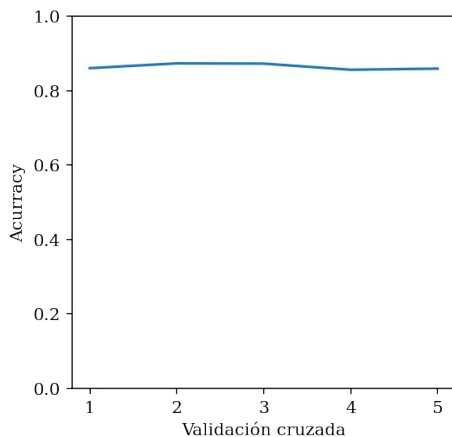


Valores de 0: 5848
Valores de 1: 650

Desbalance

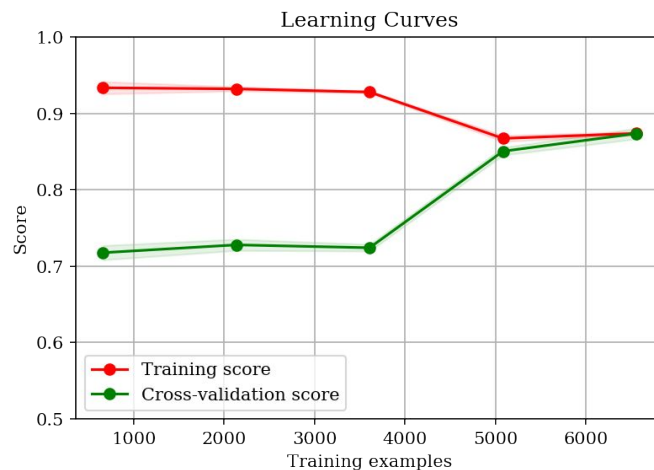


Validación cruzada del modelo lr con oversampling



Se obtiene un buen rendimiento del modelo lr con oversampling sin presentar problemas de varianza, solo un poco de bias.

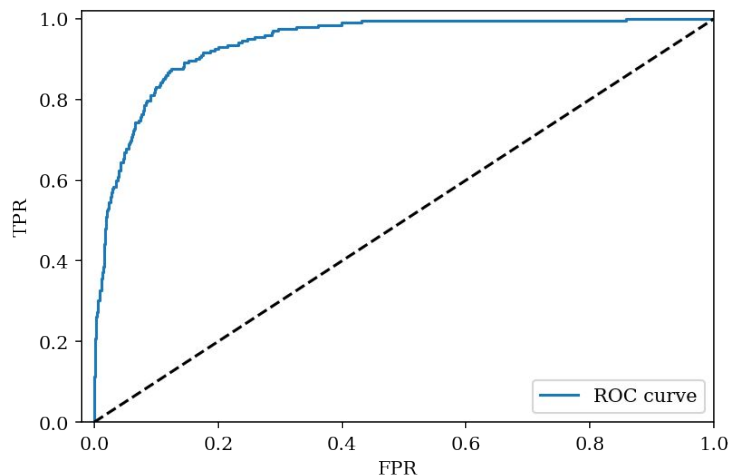
El desempeño se mantiene estable entre validaciones y se reduce el underfitting entre más muestras se tengan, al final de la curva se presenta un buen desempeño de los datos concordando la validación con entrenamiento



Datos de prueba

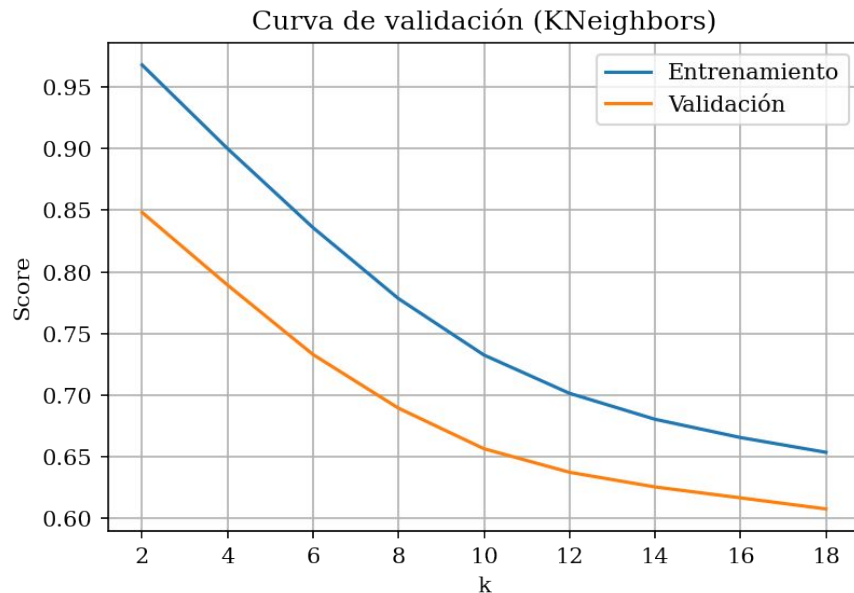
	precision	recall	f1-score	support
0.0	0.98	0.87	0.92	1748
1.0	0.43	0.86	0.58	202
accuracy			0.87	1950
macro avg	0.71	0.87	0.75	1950
weighted avg	0.93	0.87	0.89	1950

Las métricas muestran un buen desempeño en el periodo de prueba de los datos.



El umbral donde se presenta el mejor desempeño del modelo es de 0.47, muy cercano al 0.5 que toda el modelo por defecto

Prueba de hiperparametros con vecino más cercano



En este caso utilizando vecino más cercano, el mejor número de vecinos es 2, con un overfitting importante.

Gracias