

Modelo inteligente para la predicción de la tasa de abandono de los contratos de electricidad

Marcos Ornia Luna
Universidad de Oviedo

Cátedra TotalEnergies de Analítica de Datos e Inteligencia Artificial



Introducción

En el mercado eléctrico español, la gran competitividad obliga a las comercializadoras a ofrecer contratos que se ajusten cada vez más a las necesidades de los clientes y, al mismo tiempo, aseguren la sostenibilidad del negocio.

El diseño de estos contratos implica considerar una amplia variedad de variables, esta complejidad, combinada con la necesidad de mantener precios competitivos y garantizar la rentabilidad, hace crucial la comprensión del comportamiento de los clientes y la predicción de posibles bajas (*churn*).

Objetivos del estudio

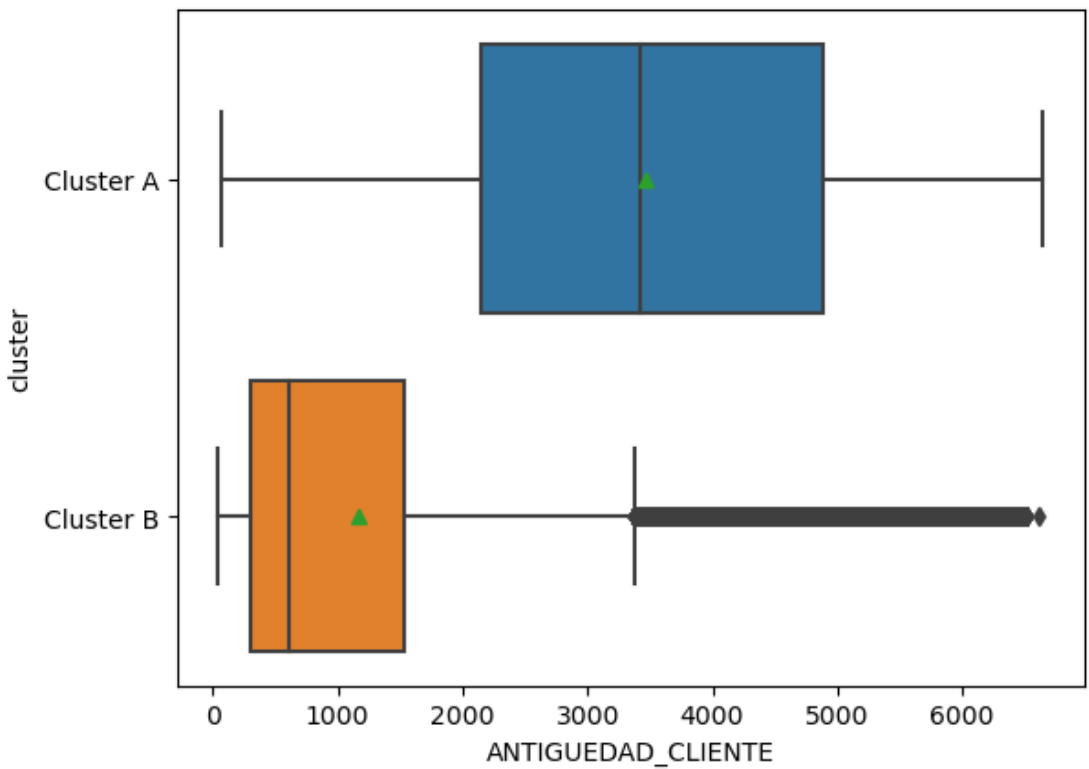
Para abordar este desafío, se desarrolló un modelo predictivo de *churn* basado en técnicas de análisis de datos y aprendizaje automático, cuyo objetivo es anticipar las bajas de clientes, entender los factores que las impulsan y proponer estrategias para minimizarlas.

Este proyecto busca optimizar los esfuerzos de retención, mejorando tanto la experiencia del cliente como los resultados empresariales.

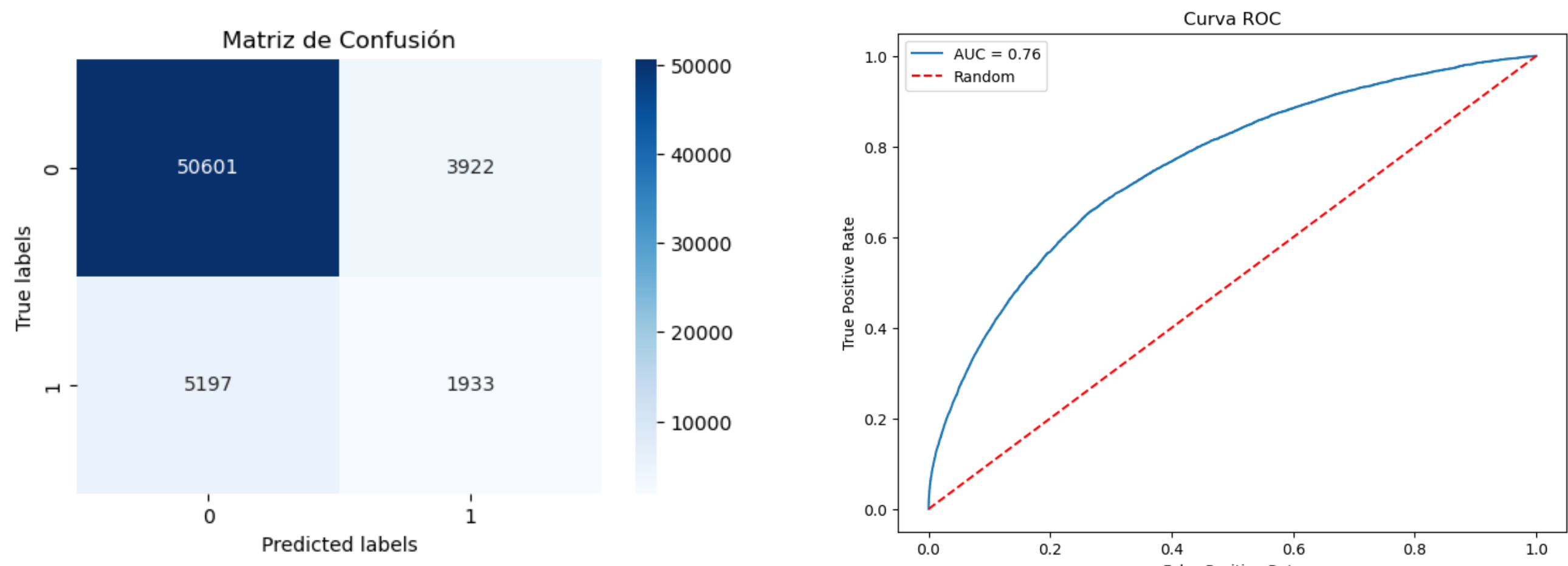
Métodos

En este proyecto se han desarrollado disitintos modelos predictivos con técnicas de inteligencia artificial explicable. Se han usado distintos enfoques como regresión logística, árboles de decision, ensambles de modelos y técnicas de clusterización.

NOMBRE_FUNCIONAL	NOMBRE_DESCRPTIVO
ANTIGUEDAD_CLIENTE	Antigüedad cliente (días)
PROVINCIA / PROVINCIA_PS	Provincia
DIAS_FIN_PROMO	Días hasta el final de la promoción
SUBCANAL_COD_INPUT / SUBCANAL_INPUT	Subcanal de venta
IMP_CONSUMO_ACTIVA_MOD	Consumo anualizado (kWh)
IMP_PRECIO_MAX	Precio
DIAS_PARA_RENOVAR_MOD	Días para renovar
N_VENTAS_CONT_FECHA_REF_3_MOD	Número de contactos tipo Ventas (últimos 3 meses)
N_PEDIDINFO_FECHA_REF_3_MOD	Número de contactos tipo Pedido de Información (últimos 3 meses)
N_CTOS_BAJA_UMES_IC_MOD	Número de contratos dados de baja (último mes)
N_CTOS_ACTIVOS_IC	Número de contratos activos del cliente
N_FACT_COBR_FECHA_REF_3_MOD	Número de contactos tipo Facturación / Cobro (últimos 3 meses)
ACCESOS_AC_ML_3MES_IC_MOD	Accesos al área de cliente en los últimos 3 meses
N_RECLAM_FECHA_REF_3_MOD	Número de reclamaciones (últimos 3 meses)

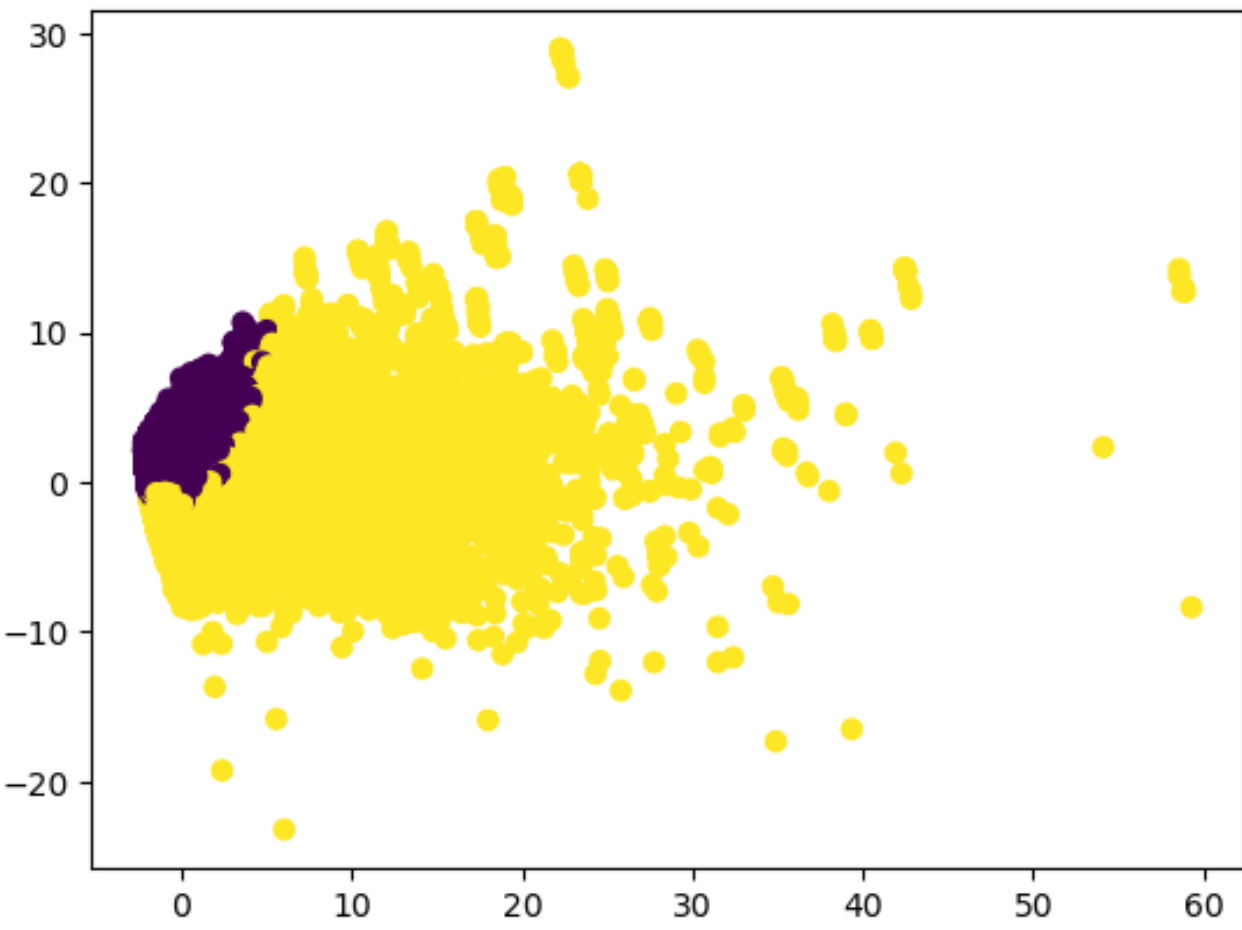
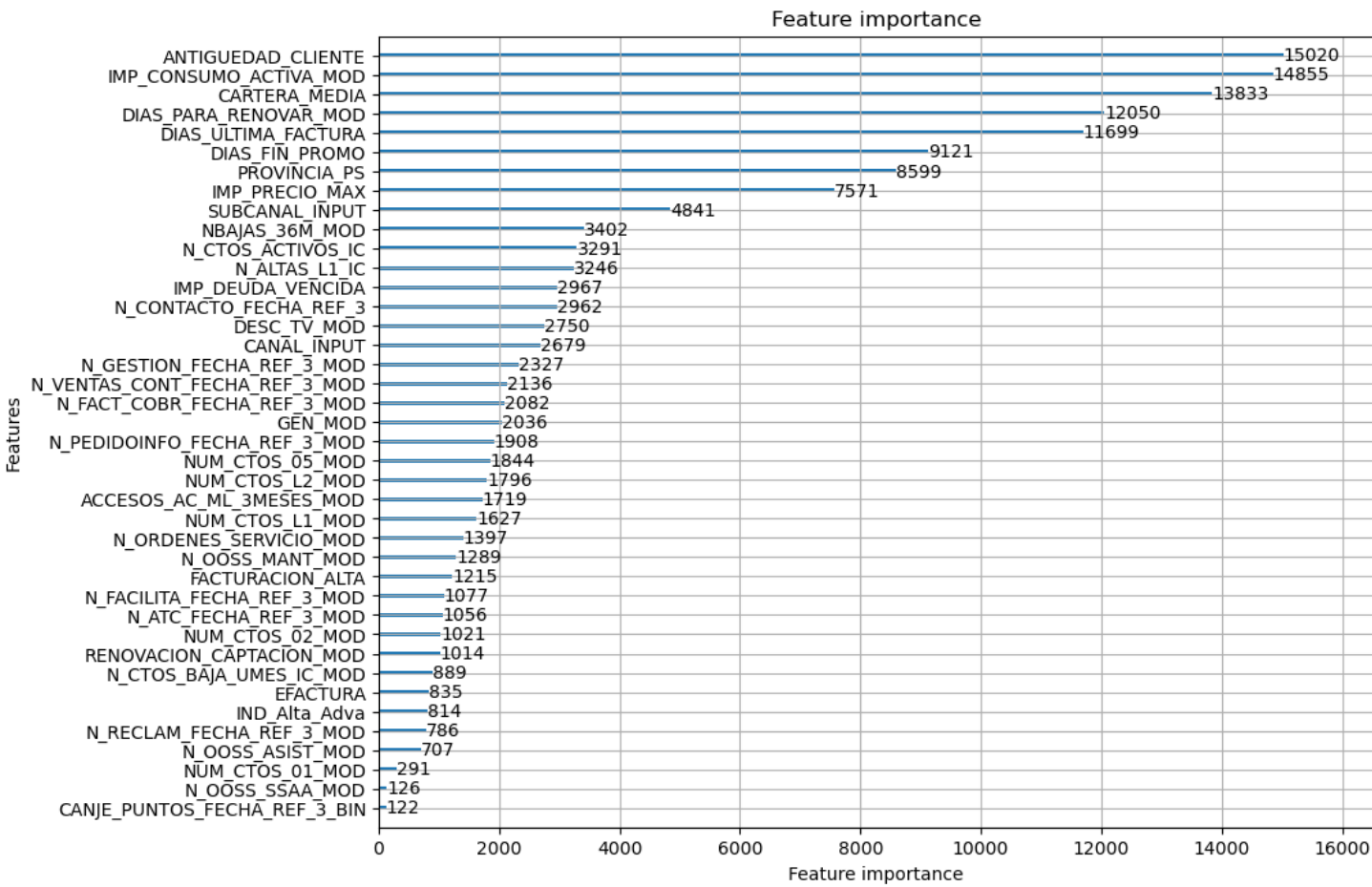


Figuras y Resultados



Métricas como la f1, la precisión o el recall dependen del umbral de clasificación. El AUC proporciona una métrica agregada del rendimiento del modelo a lo largo de diferentes umbrales de clasificación, permitiéndonos buscar un modelo equilibrado (precisión vs recall).

Para un umbral determinado la matriz de confusión nos permite observar la distribución de falsos positivos y falsos negativos, pudiendo buscar umbrales distintos en función del objetivo del modelo.



La clusterización junto con distintas técnicas explicativas, así como de visualización de la información, permitieron diferenciar claramente dos clusters donde la proporción de churn era muy distinta. Se estudio la diferencia entre estos clusters siendo muy notable en variables como la antigüedad o el tipo de contrato (anual / trimestral).

Conclusiones

Se ha podido segmentar los contratos en distintos clusters, observando como unos modelos predicen mejor la tasa de abandono en cierto tipo de contratos como son los anuales. En concreto se comprobó que los contratos trimestrales son un grupo muy homogéneo con una proporción de churn muy baja y difícil de detectar, el problema del desbalanceo de las clases se acentúa en este grupo.

Se ha explorado usar ensambles de modelos aprovechando las diferencias de rendimiento en los distintos clusters.

Se han podido ver que variables como la antigüedad, el tipo de contrato o la provincia influyen notablemente en la tasa de abandono.

El modelo que mayor rendimiento obtuvo de todos los explorados fue el LightGBM un modelo basado en boosting y árboles de decision optimizado para la velocidad y el rendimiento en grandes conjuntos de datos.

Trabajo futuro

Debido a la gran diferencia de rendimiento de los modelos en los dos grupos de contratos (anuales / trimestrales) se debería continuar buscando modelos que se adapten mejor a los modelos trimestrales, pudiendo así combinarse mediante ensambles o crear modelos separados para los contratos de un tipo y de otro.

Una futura línea de trabajo sería tratar de combinar estos modelos de machine learning con técnicas de deep learning como lo son las redes neuronales.

