1. El dataset se encuentra bastante desbalanceado y aproximadamente el 86% es de la clase 0 (no cerrará cuenta) y el 14% sí.
2. Hablando estrictamente de correlación y no de Feature Importance, las variables más correlacionadas con el Target son **reclamos\_6m** y **usa\_tarjeta\_credito.**
3. En **AUC** el mejor modelo fue **SVM** con **0.5331** y el mejor **Recall** fue **Decision Tree** con **0.1383**.
4. Dado que son “silenciosos”, se debe prestar atención a su comportamiento. Por ejemplo, se puede analizar:

* La frecuencia de uso de su cuenta (transacciones, consultas, etc.): menor frecuencia puede asociarse a una mayor probabilidad de cerrar la cuenta
* Saldo en la cuenta: reducciones en el saldo a valores muy bajos podría asociarse a una mayor probabilidad de cerrar la cuenta
* Tiempo desde la última interacción exitosa (última compra o pago): mayor tiempo podría asociarse a una mayor probabilidad de cerrar la cuenta.

1. En este caso, **falso positivo** implicaría hacerle campaña a alguien que no se va a ir mientras **falso negativo** significaría no hacerle campañas a alguien que sí se va a ir, generalmente para las empresas esto último es más grave, así que debemos evitar que nuestro modelo caiga en este error y para eso es bueno fijarnos que el F1 este bien. Ahora, algo que podemos hacer también para mitigar el problema que causan los falsos positivos, es ofrecer acciones de bajo costo para clientes que estén cerca del umbral, por ejemplo, enviarles emails.
2. El umbral optimizado por KS reduce la pérdida financiera, a comparación del umbral por default de 0.5. El modelo de predicción calcula una probabilidad de cerrar la cuenta para cada cliente, pero no recomienda un umbral a partir del cual la empresa debería realizar esfuerzos de retención del cliente. Para eso, se puede usar el test de **Kolmogorov-Smirnov** (KS), que compara las distribuciones acumuladas de los clientes que permanecen usando su cuenta vs aquellos que no, y haya el punto de corte que maximiza la discriminación entre ambas poblaciones (que es el umbral que minimiza los errores de clasificación).
3. Una posibilidad es que aunque el cliente es Premium y tiene un buen saldo, hay otras variables con valores SHAP positivos (baja satisfacción, reclamos, poca actividad digital, pocos productos, etc) qye empujan con más fuerza la predicción hacia “cerrará su cuenta” (target). Entonces aunque el Saldo empuje hacia una dirección el resto de variables podría estar empujando hacia la zona contraria.

Además, en el contexto del negocio esto no resultaría inusual: un cliente de alto nivel suele ser más exigente y esperar un servicio superior del banco; por ello, puede presentar más reclamos, menor satisfacción y eventualmente decidir cerrar su cuenta.

1. No, la tasa de interés no debería ser incluida como predictor si luego puede cambiar. Esto es porque el modelo de predicción va a estimar una probabilidad de cerrar la cuenta, que está afectada por la tasa de interés (si esta es incluida como predictor). Sin embargo, tras hallar la probabilidad, el negocio decidirá (podrá cambiar) qué tasa de interés fija para el cliente. Entonces, la tasa de interés que el modelo usó no será un dato verdadero, ya que cambió como resultado de la misma predicción. Esto puede distorsionar las estimaciones del modelo y hacerlo inestable (es decir, que al aplicarlo en producción, para predecir las probabilidades de clientes, la relación hallada por el modelo no se mantenga).
2. Los gestores humanos pueden perder confianza al ver que (a) los modelos predictivos los suplantan, y (b) desconocen qué variables toman en cuenta. Para mitigar (a): Los gestores humanos y modelos predictivos pueden trabajar en conjunto, por ejemplo, si el modelo genera una recomendación, y la decisión final la toma el gestor. En este caso, si se quiere, se puede reducir la discrecionalidad del gestor y limitarla a un rango de valores de la probabilidad hallada (por ejemplo, si el valor de la probabilidad de cerrar la cuenta se encuentra en los extremos puede primar la decisión del modelo, y si se encuentra en un rango intermedio, la decisión puede ser tomada por el gestor). Para mitigar (b): Se puede explicar a los gestores algunas de las variables más importantes que toma en cuenta el modelo, y si se relacionan positiva o negativamente con el cierre de la cuenta.