Conclusiones del Proyecto de Modelado Predictivo – TelecomX 2

Durante el desarrollo del proyecto se implementó un pipeline completo de análisis de datos,   
que incluyó limpieza, codificación, verificación, análisis exploratorio y modelado predictivo.   
Sin embargo, se identificó una limitación importante: la variable objetivo Churn solamente contiene   
una clase (0 = clientes activos) y no incluye ejemplos de cancelación (1).  
  
Esto impidió el entrenamiento de modelos de clasificación como Regresión Logística y Random Forest,   
ya que dichos algoritmos requieren la presencia de ambas clases para poder aprender y generalizar.  
  
Aun así, este hallazgo es relevante y aporta las siguientes conclusiones:  
  
1. Calidad de los datos: antes de aplicar modelos predictivos, es indispensable verificar que la variable   
 objetivo tenga representadas todas las clases necesarias.  
2. Limitación del dataset: la ausencia de clientes cancelados en los datos indica que el dataset no está   
 completo o fue procesado de manera incorrecta.  
3. Imposibilidad de análisis de importancia de variables: sin un modelo entrenado no es posible identificar   
 los factores más influyentes en la cancelación.  
  
Recomendaciones para proyectos futuros:  
  
- Revisar y corregir el dataset original para asegurar que Churn contenga tanto clientes activos (0)   
 como cancelados (1).  
- En caso de que existan muy pocos cancelados, aplicar técnicas de balanceo de clases (SMOTE, sobremuestreo   
 o ponderación de clases en los modelos).  
- Reentrenar los modelos con los datos corregidos, lo cual permitirá:  
 - Evaluar métricas de desempeño (accuracy, precision, recall, F1-score).  
 - Analizar la importancia de variables (coeficientes en Regresión Logística, importancia en Random Forest).  
 - Proponer estrategias de retención basadas en los factores que más influyen en la cancelación.  
  
Reflexión final:  
  
Aunque los modelos no pudieron entrenarse por falta de datos representativos, el pipeline implementado demuestra   
el proceso completo de un proyecto de analítica de datos y machine learning. Además, el resultado obtenido resalta   
la importancia crítica de contar con datos de calidad y balanceados para el éxito de proyectos predictivos.