

Sustentación del Modelo deserción de clientes.

Resumen de Pasos:

1. **Configuración del Ambiente:** Se configuró el entorno importando las bibliotecas necesarias y cargando el conjunto de datos desde un archivo CSV, obtenido mediante un enlace proporcionado. Se realizó una limpieza inicial del archivo para eliminar caracteres no deseados (""") y asegurar la integridad en la lectura del mismo.
2. **Exploración de Datos:** Se aplicaron técnicas avanzadas de Análisis Exploratorio de Datos (EDA) para profundizar en la comprensión de la información. Se realizaron análisis de estadísticas descriptivas para todas las variables, identificando valores atípicos y distribuciones no normales que podrían afectar el rendimiento del modelo. Se generaron matrices de correlación para explorar las relaciones entre las variables, utilizando gráficos de calor para visualizar las fuerzas y direcciones de estas correlaciones. Además, se realizaron análisis de frecuencias y distribuciones de variables categóricas y numéricas para identificar patrones significativos y características relevantes que podrían influir en la predicción de la deserción de clientes.
3. **Procesamiento de Datos:** Se procedió al preprocesamiento exhaustivo de los datos para prepararlos adecuadamente para la modelización. Se implementaron técnicas de ingeniería de características, como la definición de rangos para variables numéricas discretas, la consolidación de columnas con alta correlación para reducir la multicolinealidad y mejorar la eficiencia del modelo, y la transformación de variables categóricas para su integración en algoritmos de aprendizaje automático. Asimismo, se realizó la normalización de los datos para estandarizar las escalas y se aplicó un proceso de balanceo de clases para mitigar el desequilibrio en la distribución de la variable objetivo, asegurando así que el modelo no esté sesgado hacia la clase mayoritaria y pueda generalizar mejor a datos nuevos.
4. **Construcción de Modelos:** Se exploraron y compararon varios modelos de aprendizaje automático para predecir la cancelación de productos bancarios por parte de los clientes. Se evaluaron modelos de clasificación como Regresión Logística, Árboles de Decisión, Random Forest y Naive Bayes. Cada modelo fue entrenado utilizando técnicas de validación cruzada para optimizar los hiperparámetros y mejorar la capacidad de generalización. Se implementaron métricas de evaluación como precisión, recall, F1-score y área bajo la curva ROC (AUC-ROC) para medir el rendimiento de cada modelo en el conjunto de prueba. Finalmente, se seleccionó el modelo de Random Forest debido a su consistente rendimiento superior al 90% en todas las métricas evaluadas, demostrando su capacidad para identificar de manera efectiva los clientes propensos a la deserción de productos bancarios.

5. **Preparación del Dataframe de Salida para Power BI:** Se realizaron ajustes adicionales al dataframe, incluyendo la definición de rangos en variables específicas, con el fin de facilitar su interpretación en el dashboard final.
6. **Dashboard en Power BI:** Utilizando el dataframe preparado, se procedió a importar los datos y crear un dashboard que presenta gráficos categorizados para proporcionar una comprensión profunda de los datos.

Este diagrama y resumen condensan de manera concisa cada uno de los pasos realizados en el notebook para abordar el problema de negocio planteado por el banco sobre la deserción de clientes.

