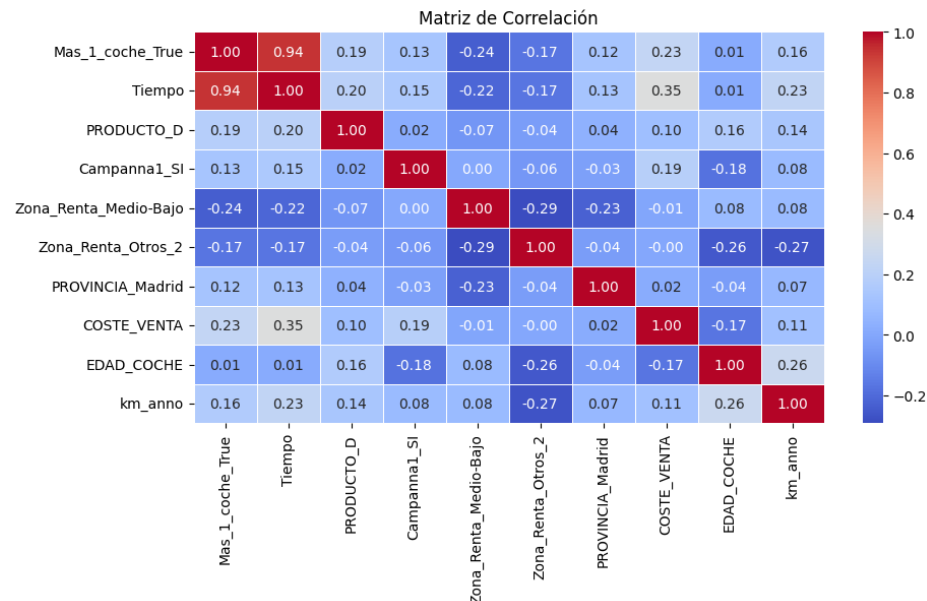
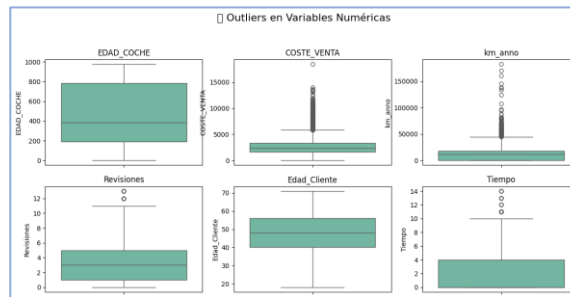
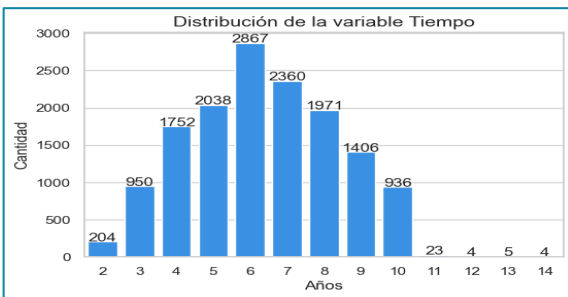
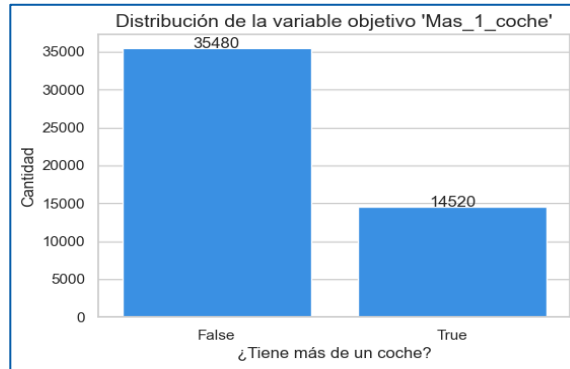
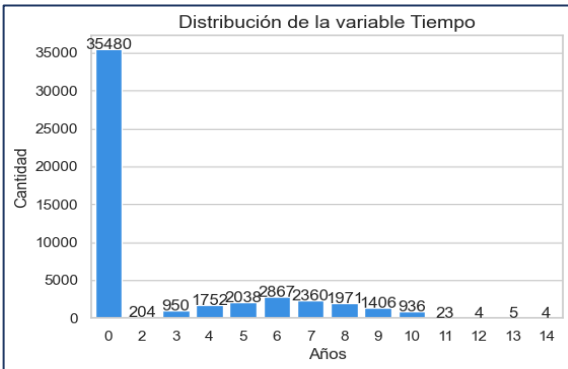


Análisis exploratorio de los Datos



DESCUBRIMIENTOS

Nulos: Estado Civil - Genero - Zona Renta

Datos bien tipificados

Los valores atípicos concuerdan con el contexto del negocio

Desbalance en las Variables objetivos (Mas_1_Coche, Tiempo)

PERFIL DEL CLIENTE PROPENSO A COMPRAR - EDA

Los que han comprado el Producto_D

Renta Alta (Correlación negativa con Renta Media Baja)

Mayoritariamente de la provincia de Madrid

Han efectuado más la Campaña 1 (Descuento por financiación)

Cierta correlación positiva con el kilometraje

A mayor coste de venta, más se compra un 2n Vehículo.

SkLearn Pipeline

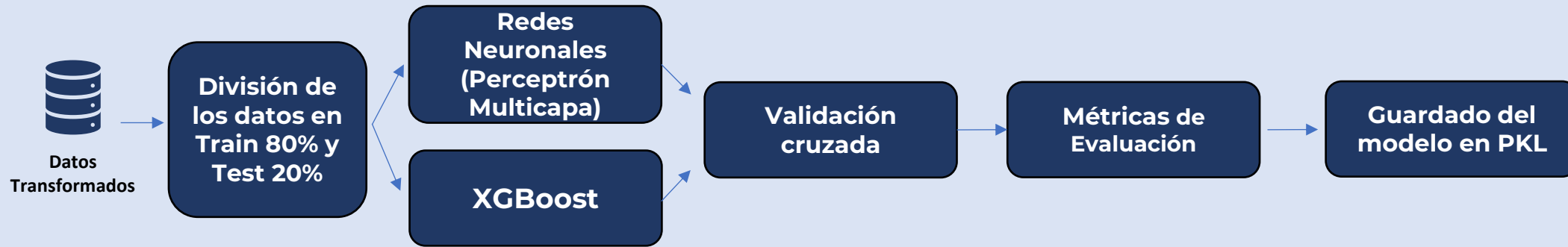
Eliminamos columnas Irrelevantes (CODE)

Reemplazo de Valores Nulos en Columnas Categóricas

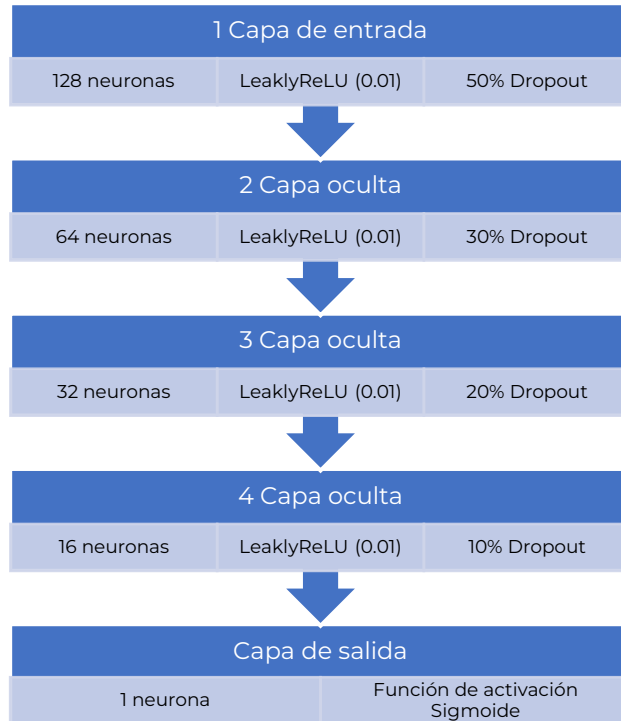
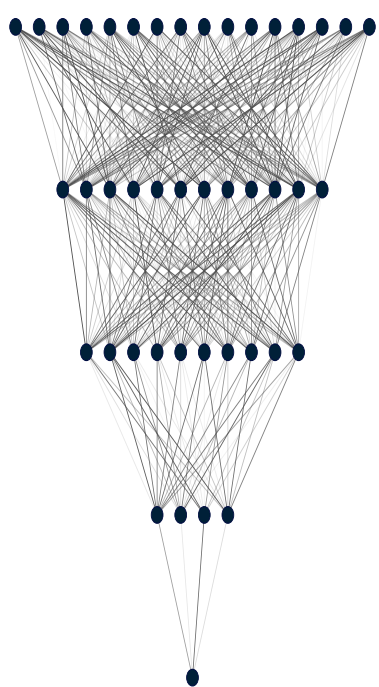
Codificación OHE de Variables Categóricas, Booleanas

Escalado de Variables Numéricas (Standard Scaler)

Modelo de clasificación y regresión



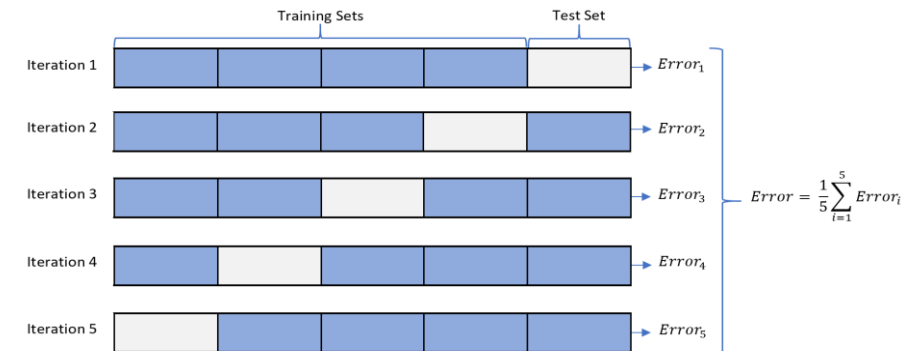
Arquitectura del MLP



XGBoost

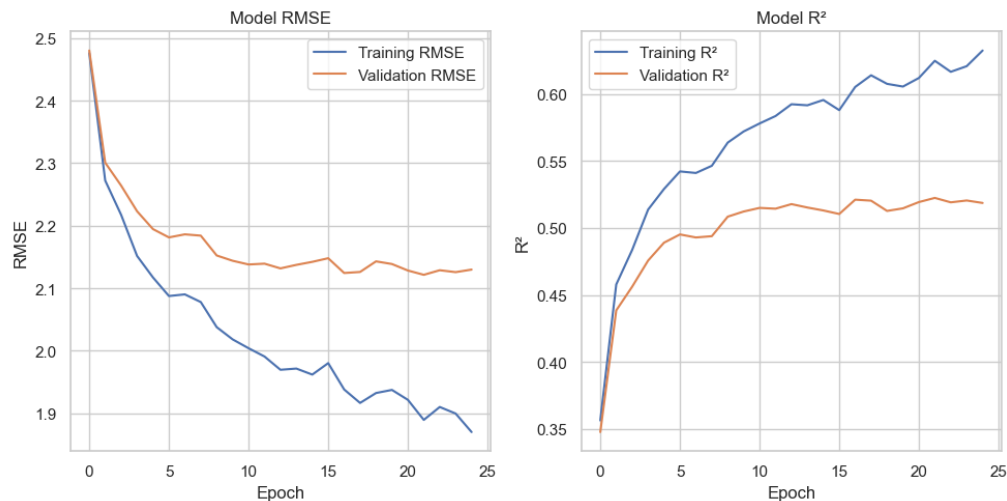
- Número de Árboles:** 1000 árboles
- Profundidad Máxima:** de 6 niveles.
- Early stopping:**
- Tasa de aprendizaje** (0.01)

Validación cruzada y Métricas utilizadas



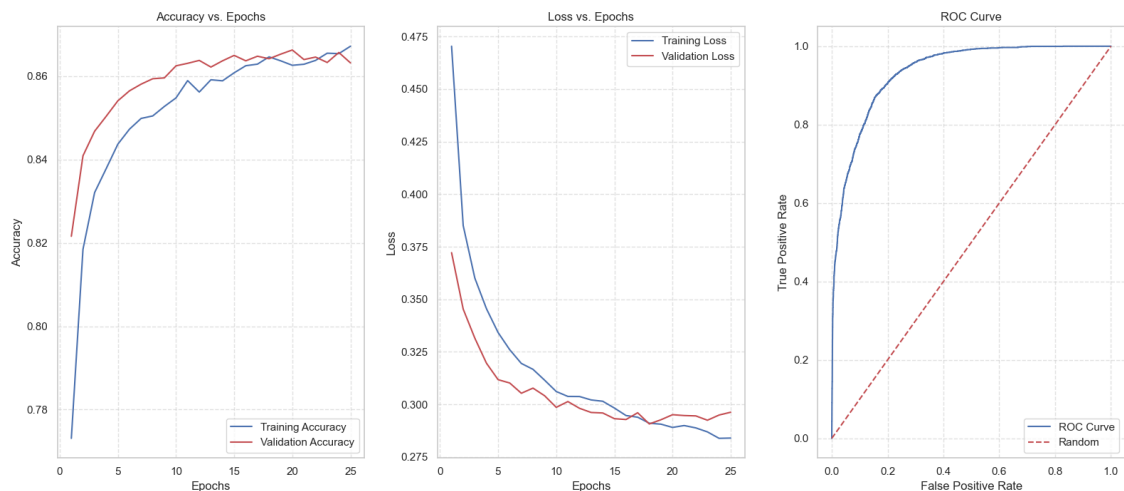
CLASIFICACIÓN: Matriz de confusión, Accuracy, Recall, F1 Score, Curva ROC.

REGRESIÓN: RMSE



Modelo de Regresión con Redes Neuronales

- El **R²** mide la proporción de la varianza en la variable objetivo explicada por el modelo. Un valor de 0.8743 (o 87.43%) indica que el modelo captura una gran parte de la variabilidad en el tiempo entre compras, lo cual es un resultado excelente.
- El **RMSE**, que es la raíz cuadrada del MSE (0.7073), se expresa en la misma unidad que la variable objetivo (años). Esto significa que, en promedio, las predicciones del modelo se desvían aproximadamente 0.71 años de los valores reales. Comparado con la media (6.3624 años) y la desviación estándar (1.9995 años), este error es significativamente menor que la variabilidad natural de los datos, lo que sugiere una buena precisión del modelo.



La exactitud de entrenamiento (azul) y validación (rojo) aumenta rápidamente en las primeras 5-10 épocas, estabilizándose alrededor de 0.85-0.86. La cercanía de las curvas indica que el modelo generaliza bien sin sobreajuste significativo.

La pérdida de entrenamiento y validación disminuye desde un valor inicial de 0.45 a aproximadamente 0.25-0.30 tras 25 épocas. La convergencia de ambas curvas sugiere un aprendizaje estable, con una ligera fluctuación en la validación que podría indicar pequeños ajustes necesarios.

La curva ROC (azul) muestra un área bajo la curva de 0.8198, superando la línea diagonal de referencia (roja), lo que confirma una buena capacidad de discriminación. La curva se eleva rápidamente, indicando un buen equilibrio entre tasa de verdaderos positivos y falsos positivos.

Modelo de Clasificación con Redes Neuronales

- El **Accuracy** mide la proporción de predicciones correctas sobre el total de predicciones. Con un valor de 0.8642 (86.42%), el modelo acierta en más de las cuatro quintas partes de los casos, lo que indica un rendimiento sólido. Esto se alinea con los resultados de validación cruzada (0.8631 ± 0.0021), confirmando la consistencia del modelo.
- El **recall** mide la capacidad del modelo para identificar correctamente a los clientes que harán una segunda compra ($TP / (TP + FN)$). Un valor de 0.7138 (71.38%) indica que el modelo detecta más de dos tercios de los casos positivos, lo que es aceptable, aunque deja un 28.62% de falsos negativos. Esto es consistente con la validación cruzada (0.7181 ± 0.0231), mostrando estabilidad.
- El **F1 score** es el promedio armónico de precisión y recall, ofreciendo un balance entre ambos. Un valor de 0.7533 refleja un compromiso razonable entre identificar correctamente a los positivos y minimizar falsos positivos, aunque podría mejorarse con un ajuste en el umbral.
- El **área bajo la curva ROC** mide la capacidad del modelo para distinguir entre clases positivas y negativas. Un valor de 0.8198 (81.98%) indica una buena discriminación, superior al valor aleatorio (0.5), aunque está por debajo del resultado de validación cruzada (0.9351 ± 0.0010), lo que podría sugerir una ligera sobreestimación en la validación cruzada o diferencias en los datos.
- La **validación cruzada** con 5 pliegues muestra una precisión promedio de 0.8631 (± 0.0021), una precisión de 0.7918 (± 0.0112), un recall de 0.7181 (± 0.0231), y un ROC-AUC de 0.9351 (± 0.0010). Estos valores son consistentes con las métricas individuales, con una variabilidad baja (desviaciones estándar pequeñas), lo que respalda la robustez del modelo.

Solución de Negocio para Incentivar la Segunda Compra y Acortar el Tiempo

Modelos Utilizados

- **Modelo de Clasificación:** Predice probabilidad de segunda compra (Exactitud: 86.42%, ROC-AUC: 0.8198).
- **Modelo de Regresión:** Estima tiempo entre compras (R^2 : 0.8743, RMSE: 0.7073 años).

Estrategia: Segmentación y Acciones Personalizadas

- **Paso 1:** Identificación con el Modelo de Clasificación
 - **Grupo A:** Alta Probabilidad de Compra (> 0.5): Acción: Incentivos inmediatos (descuento 10% en 90 días, financiamiento preferencial).
 - **Grupo B:** Baja Probabilidad de Compra (≤ 0.5): Campañas agresivas (pruebas gratuitas, eventos exclusivos, programas de fidelidad).
- **Paso 2:** Predicción y Reducción del Tiempo con el Modelo de Regresión
 - **Subgrupo A1: Tiempo Corto (≤ 2 años)** Acción: Reforzar decisión con descuentos inmediatos.
 - **Subgrupo A2: Tiempo Largo (> 2 años)** Acción: Planes de renovación anticipada (intercambio con valor adicional en 6 meses), notificaciones personalizadas.
- **Paso 3:** Personalización con Feature Importance
 - Identificar factores clave (ej. PROVINCIA_Madrid) para campañas locales (eventos en Madrid).

Implementación de Incentivos

- **Programas de Lealtad:** Puntos acumulables para descuentos o accesorios.
- **Campañas Temporales:** Promociones limitadas (ej. 15% descuento antes de fin de año).
- **Notificaciones Personalizadas:** Recordatorios basados en tiempo predicho (ej. incentivo a 2.5 años si predice 3 años).
- **Financiamiento Atractivo:** Planes flexibles con tasas bajas para Subgrupo A2.

Monitoreo y Ajuste

- **Seguimiento de Resultados:** Comparar tasa y tiempo de compras antes/después de incentivos.
- **Ajuste del Umbral:** Bajar a 0.4 para aumentar recall si se priorizan más clientes.
- **Optimización del Modelo:** Explorar transformaciones (ej. logarítmica) si MAPE (40.37%) afecta predicciones.

Thank you!
¡Gracias!

Soy Innovación
Soy Maker
Soy UAX