

Executive Summary del Proyecto

Este proyecto tuvo como objetivo fundamental estimar dos métricas clave en el ámbito del marketing: **Customer Lifetime Value (LTV)** y **Customer Acquisition Cost (CAC)**. Estas métricas son cruciales para las decisiones estratégicas de marketing, ya que permiten calcular el retorno sobre la inversión en marketing (**ROMI**) y guiar la asignación de presupuestos a las fuentes más rentables.

En la fase de modelado, se entrenaron y evaluaron un total de cinco modelos, incluidos: Lineal, Ridge, Lasso, SGD y LightGBM. Todos los modelos fueron ajustados y evaluados con el **RMSE (Root Mean Squared Error)** como métrica principal de desempeño, la cual permite evaluar la precisión de las predicciones. Después de realizar la comparación, se seleccionaron los modelos más adecuados para cada métrica:

- **Ridge** se destacó como el mejor modelo para predecir LTV debido a su capacidad para manejar datos de manera estable y su rendimiento superior en comparación con los demás.
- **LightGBM** fue el modelo más adecuado para CAC, destacándose por su capacidad de manejar relaciones no lineales complejas en los datos.

Es importante señalar que, debido a limitaciones de tiempo y carga de trabajo, no se implementó validación cruzada con `TimeSeriesSplit`, ni se utilizaron técnicas avanzadas de ajuste de hiperparámetros, como el uso de `GridSearchCV`. Sin embargo, la selección del mejor modelo se realizó de forma lógica, basándose en los resultados empíricos obtenidos a través del entrenamiento de los modelos.

En cuanto a la interpretabilidad de los modelos, se generaron gráficos utilizando **SHAP** (Shapley Additive Explanations) para entender el impacto de las variables sobre las predicciones del modelo. Se construyeron dos tipos de visualizaciones:

- `SHAP Value.png` y `SHAP Value (Impact on Model Output).png`, que ayudaron a analizar cómo influyen las variables en la predicción de LTV y CAC.
- `Residuals for LTV Model.png` y `Residuals for CAC Model.png`, que proporcionaron un diagnóstico de los errores en las predicciones.

Aunque se logró una interpretabilidad razonable, algunas técnicas más avanzadas, como los gráficos de **Partial Dependence Plots (PDP)** y el análisis de errores sistemáticos por segmentos específicos, no se implementaron en esta fase del proyecto debido a limitaciones de tiempo y recursos. A pesar de esto, se

priorizó la generación de salidas clave que permitieran interpretar de manera clara cómo las variables clave influyen en las predicciones.

Respecto a la simulación de la estrategia de marketing, se construyó una función llamada `simulate_romi_from_real_values` que permitió calcular el **ROMI base** y simular varios escenarios de reasignación de presupuesto en marketing. Se consideraron dos escenarios principales:

1. Reasignar un **+10% a la mejor fuente de adquisición**.
2. Redistribuir el presupuesto de manera **proporcional según el ROMI histórico**.

Sin embargo, la ejecución final de la simulación utilizando la función `simulate_marketing_budget()` no se completó debido a un error en uno de los parámetros requeridos (falta del parámetro `source_assignment`), lo que impidió que la simulación final se llevara a cabo. A pesar de esto, se logró ejecutar una simulación preliminar que mostró las tendencias y permitió obtener algunas recomendaciones.

Conclusión General:

El proyecto logró avanzar significativamente en el análisis y la predicción de métricas clave para la toma de decisiones de marketing, como LTV, CAC y ROMI. A pesar de las limitaciones de tiempo y recursos, se sentaron las bases para futuras iteraciones del modelo, con espacio para la optimización y la implementación de técnicas avanzadas de validación y diagnóstico. Se generaron visualizaciones útiles que facilitaron la comprensión del impacto de las variables y los errores del modelo, y se dejaron listas las funciones para realizar simulaciones más robustas en el futuro.