



Camino hacia la Retención de Clientes en el Sector Financiero

Este es un desafío habitual que enfrentan las empresas financieras, la capacidad de predecir y comprender los factores que contribuyen al abandono de los clientes puede marcar una gran diferencia en la **rentabilidad** y el **éxito** de una entidad financiera.

Descubriendo Oportunidades

Exploraremos cómo el DA y el ML pueden brindar una ventaja competitiva al identificar **patrones** y **señales** tempranas para abordar un problema.

Al comprender las razones detrás de la fuga de clientes, las instituciones financieras pueden implementar **estrategias efectivas de retención**, adaptar sus **productos y servicios**, y brindar una **experiencia personalizada** para fomentar la lealtad del cliente.

Para esto, necesitamos un sistema

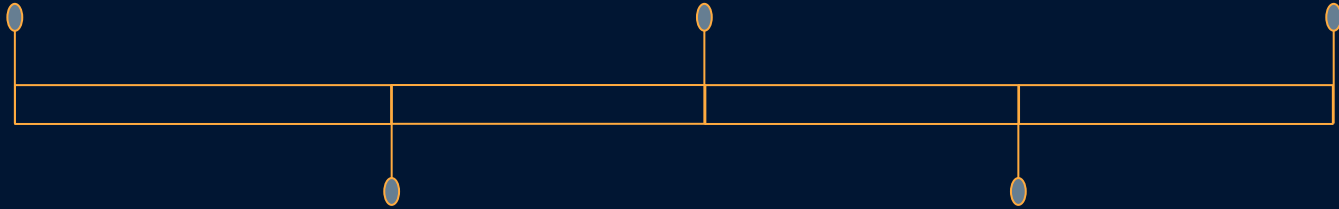


Sistema

Extracción e importación
Data Acquisition

Preparación de datos
Data Wrangling

Descubrimiento
Insights



EDA
Análisis exploratorio de los datos

Algoritmos de Machine Learning
Entrenamiento y predicción



Data Acquisition



Este es el proceso de recolectar y obtener información relevante para su posterior análisis.

Para lograrlo se utilizó



python™



NumPy



pandas

EDA

Análisis exploratorio de los datos

Este proceso implica explorar y examinar los datos con el fin de comprender su estructura, identificar patrones y tendencias, es nuestra **familiarización** con los datos

¿En qué consiste?

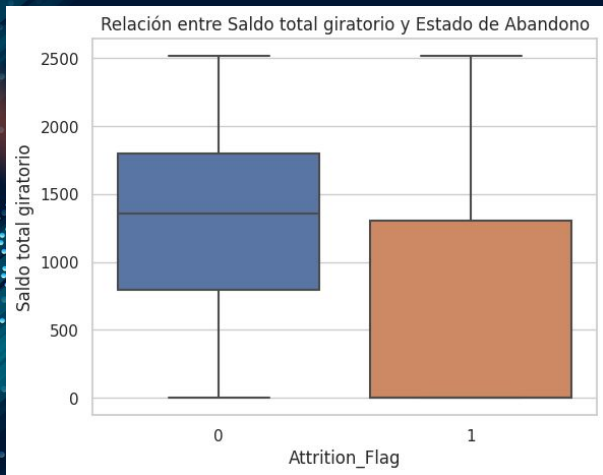


EDA

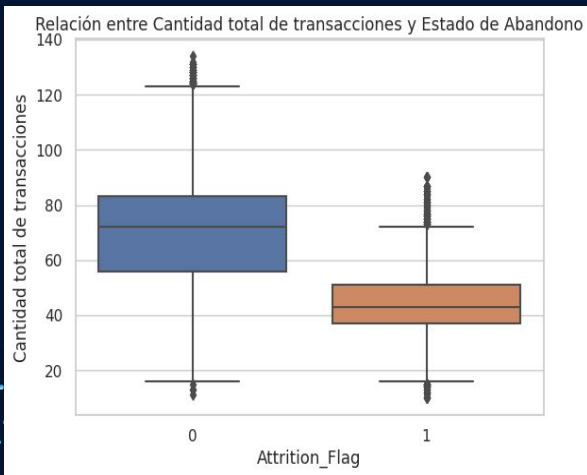
Variables en función de nuestra variable target

Análisis exploratorio de los datos

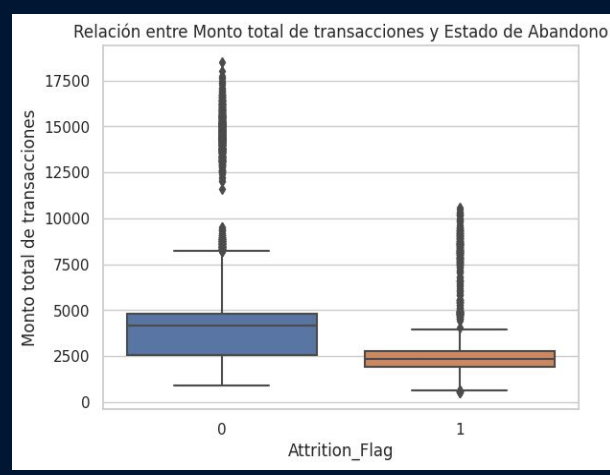
Saldo total Giratorio



Cantidad de transacciones



Monto de transacciones



EDA

Análisis exploratorio de los datos

Si un cliente hace una **cantidad total** de transacciones de lo normal.

¿Hay más probabilidad de abandono?



Si el cliente **no utiliza el crédito** disponible regularmente.

Si un cliente hace un **giro de fondos** más alto de su promedio.



Data Wrangling

Este es un proceso de preparación de los datos posterior a la implementación de algoritmos

¿En qué consiste?

- **Alcance:** Determinar los objetivos y los resultados esperados en función de los datos disponibles.
- **Estandarización de datos:** Transformar los datos a una escala común para facilitar su comparación y análisis.
- **División de datos y PCA:** Separar los datos en conjuntos de entrenamiento (train) y prueba (test) ayudará a evaluar realmente la capacidad predictiva de un modelo.

Algoritmos de Machine Learning

En esta etapa, tenemos entendido que nuestro problema es de clasificación binaria.

¿Cómo es el proceso?



Modelos de Machine Learning Implementados

1. **Random Forest** Combina múltiples árboles de decisión para obtener un resultado final.
2. **Decision Trees:** Árbol jerárquico que toma decisiones basadas en características para predecir resultados.
3. **Gaussian Naive Bayes:** Calcula la probabilidad de un evento basándose en eventos relacionados utilizando el Teorema de Bayes.
4. **LGBMClassifier:** Modelo basado en Gradient Boosting que utiliza árboles de decisión para mejorar la precisión gradualmente.



¿Qué modelos demostraron mejor desempeño?

Para abordar esta interrogante, nos basaremos en las métricas obtenidas por cada algoritmo, considerando indicadores como la **Accuracy**, **Precision**, **Re-Call** y **F1-Score**.

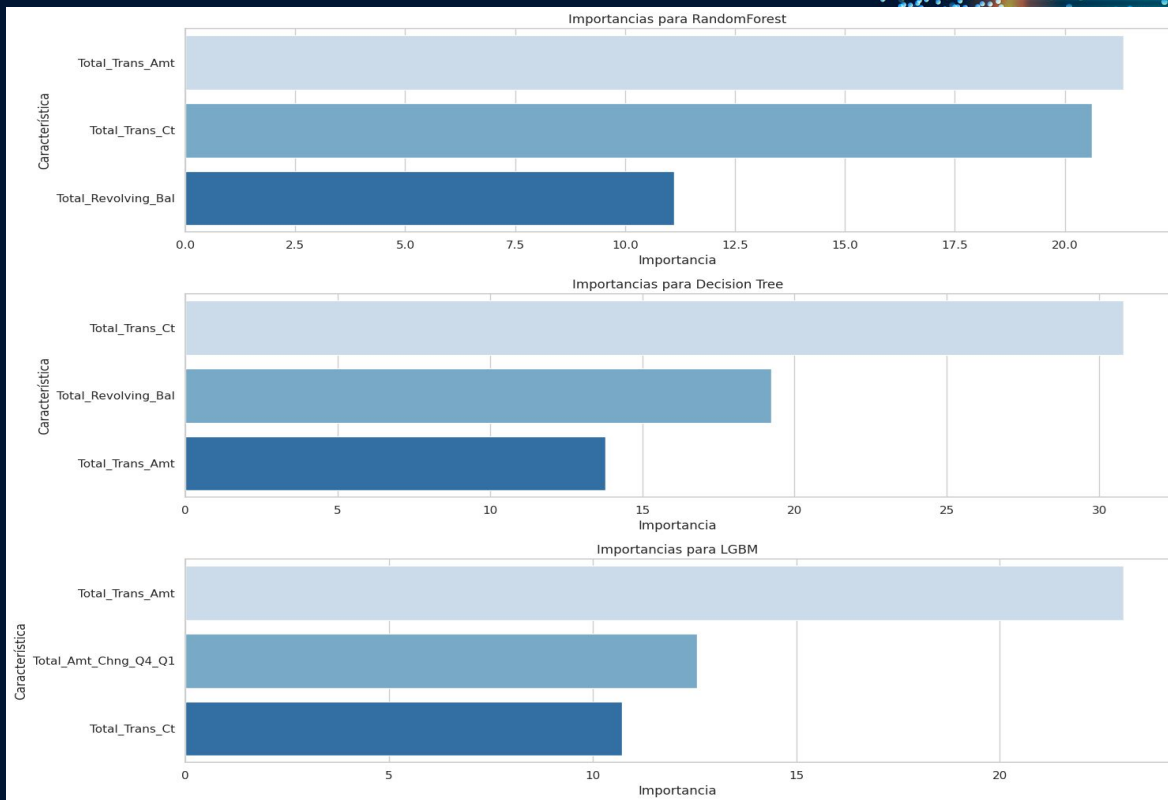
Para alcanzar este objetivo, es necesario respondernos las siguientes interrogantes

- ¿Cuál es la Importancia de las variables en los modelos?
- ¿Qué modelo presentó mejor accuracy?
- ¿Qué modelo se desempeñó mejor en todas las métricas?
- ¿Cuáles fueron los modelos que mostraron menos diferencia entre las métricas test y train?
- ¿Cómo se distribuyen los valores de las métricas?

¿Cuál es la Importancia de las variables en los modelos?

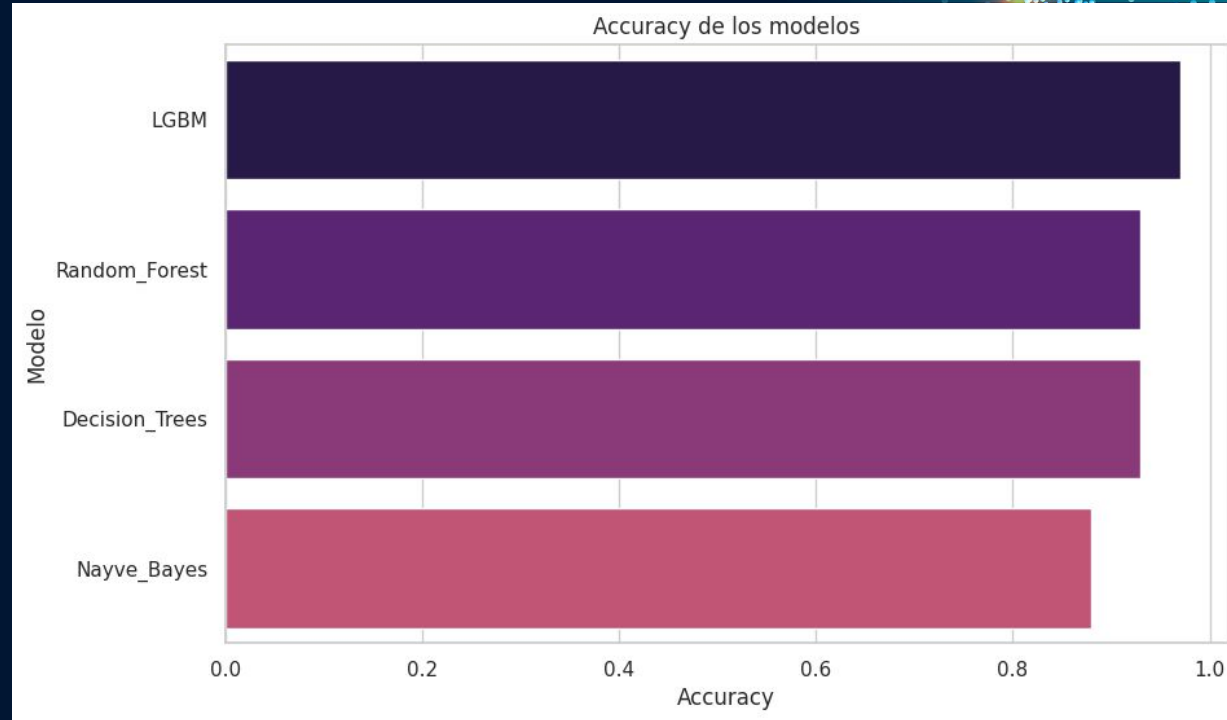
En este gráfico podemos entender cuáles fueron las columnas más influenciadas para cada modelo.

Total_Trans_Ct, **Total_Trans_amt** y **Total_Revolving_Bal** fueron las más **predominantes**.

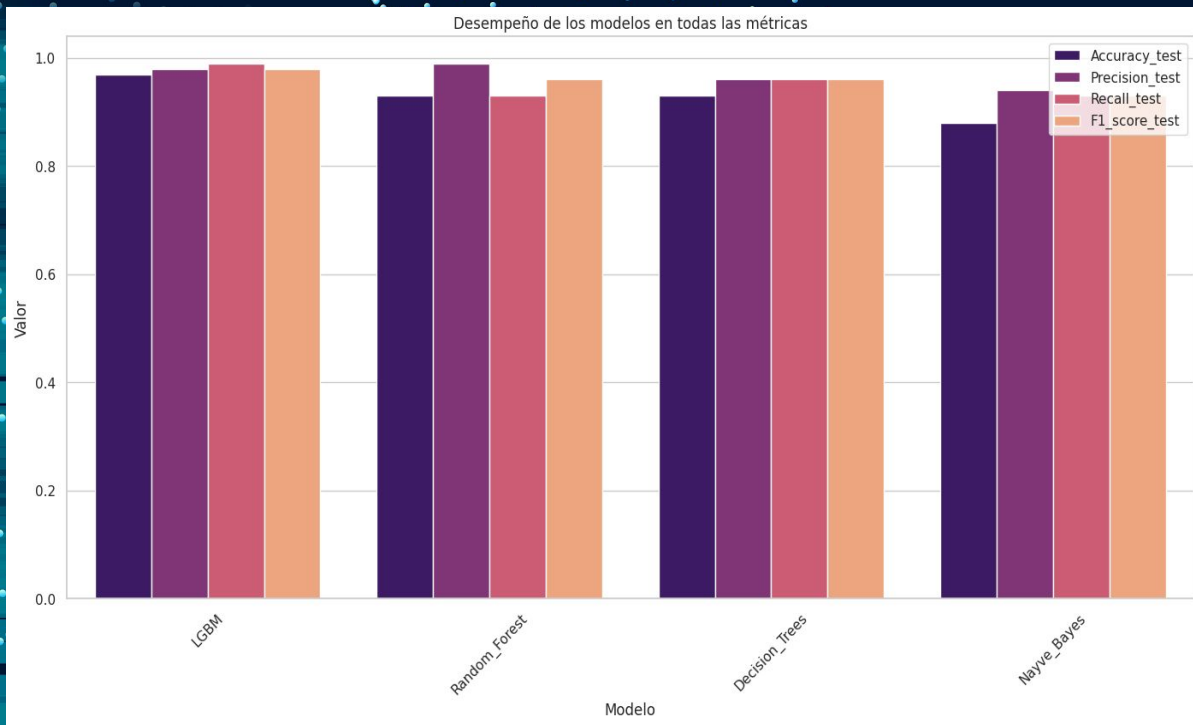


¿Qué modelo presentó un mejor Accuracy?

LGBM presentó un mejor accuracy a diferencia del resto de los modelos. Sin embargo, no es una diferencia considerable.

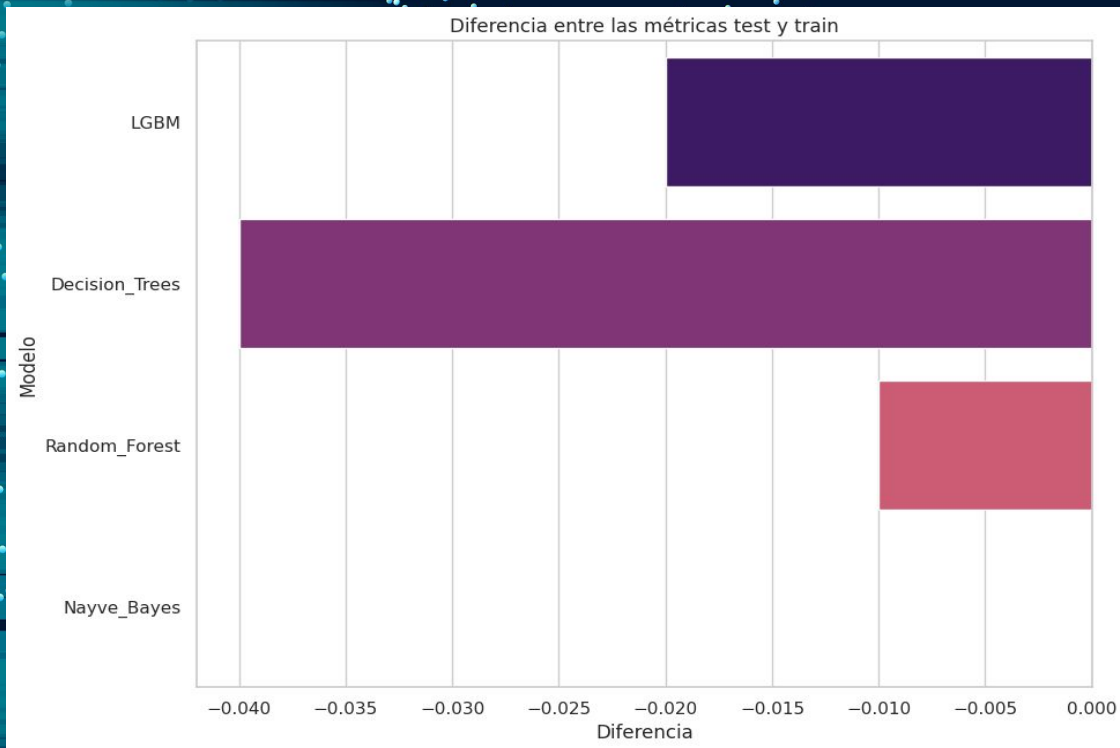


¿Qué modelo mostró un rendimiento superior en todas las métricas?



Se puede comprender que el modelo **LGBM** presentó mejor desempeño en todas las métricas desarrolladas. Siendo **Naive_Bayes** el modelo con menor desempeño.

¿Cuáles fueron los modelos que mostraron menos diferencia entre las métricas test y train?



Una particularidad descubierta es que **Nayve_Bayes** no presenta ninguna fluctuación entre los modelos de **train** y **test**. Siendo el modelo Decision_Trees el más afectado, este presenta mayor **diferencia** entre los valores de métricas train y test.

¿Cómo se distribuyen los valores de las métricas?

Podemos identificar la distribución de valores de las métricas en todos los modelos.

Siendo el **accuracy** con valores más diferentes.

La **precisión** presenta menos volatilidad.



Conclusiones

En este proyecto, abordamos el desafío del abandono de clientes, buscando descifrar las causas detrás de cada pérdida. Tras análisis y experimentación, seleccionamos **LightGBM como el modelo idóneo**. Destacó en todas sus métricas como en su equilibrio en datos de entrenamiento y prueba, sin distancia significativa entre clases .

Damos como concluido este proyecto comprendiendo las **causas subyacentes** del abandono, **alertando posibles casos** y recomendando aplicar estrategias de retención y personalización para aquellos clientes.

Gracias por tu tiempo

Te invito a revisar el proyecto
desde GitHub

Enlace

Cualquier feedback es bienvenido