

Detección de Fraude Financiero

Data Scientists:



Gerardo



Cristobal



Isaias



Análisis del Problema

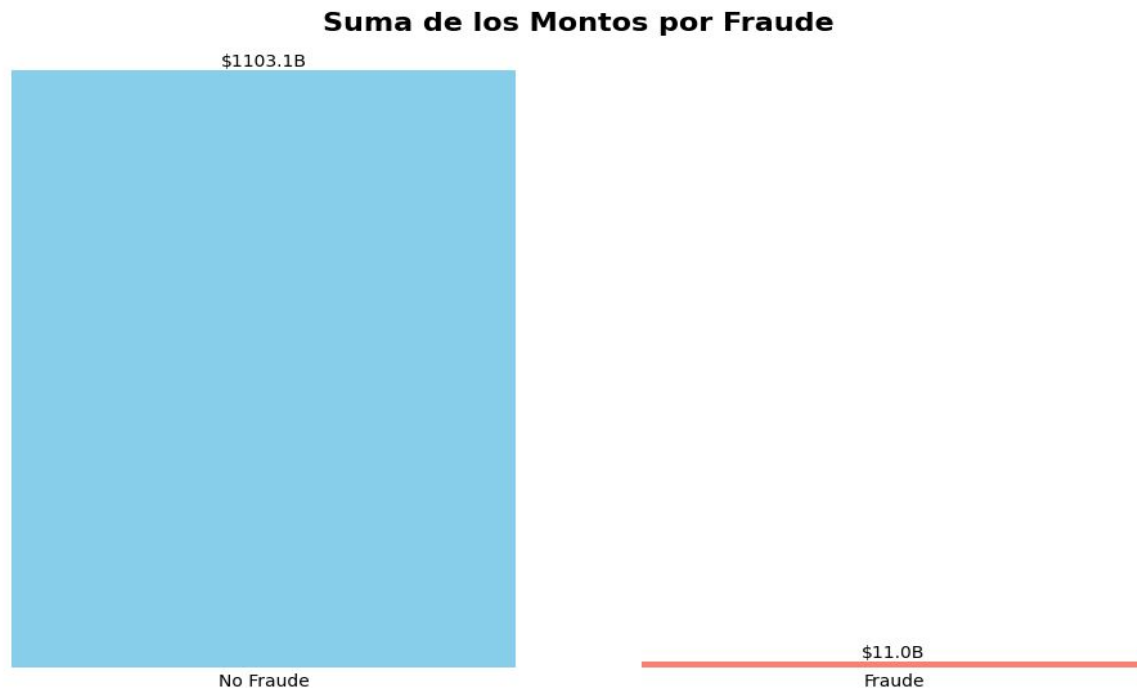
La urgencia por detectar fraudes en transacciones móviles de dinero ha llevado a una empresa del segmento Fintech a buscar soluciones innovadoras.

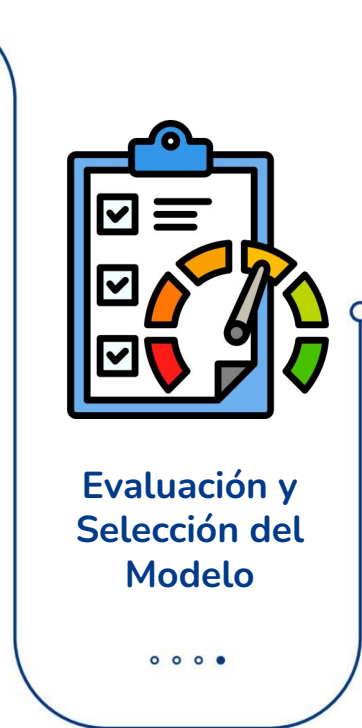
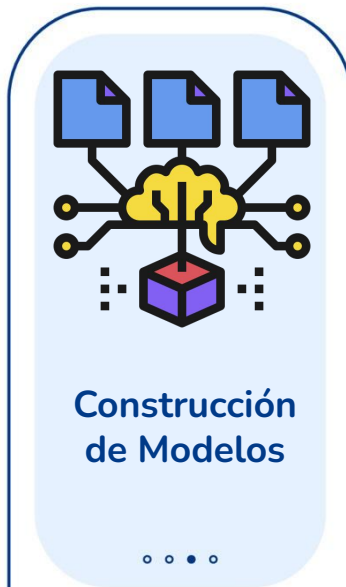
Este proyecto utiliza técnicas de Machine Learning para detectar transacciones fraudulentas con el objetivo de reducir las pérdidas financieras y mejorar la seguridad de las transacciones.



Según un informe de la Asociación de Profesionales Certificados en Fraude (ACFE), las organizaciones **pierden** aproximadamente un **5%** de sus ingresos anuales debido a fraudes.

Tan sólo el **último mes**, los montos por fraude ascendieron hasta los **\$11B**.







Preprocesamiento de Datos

Se realizó la limpieza de datos:

- ★ Creación de una nueva variable
- ★ One Hot Encoding
- ★ Eliminación de columnas prescindibles
- ★ Eliminación de duplicados
- ★ Tratamiento de nulos

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

step		type	amount	nameOrig	oldbalanceOrg	newbalanceOrg	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
0	1	PAYMENT	9839.64	C1231006815	170136.0	160296.36	M1979787155	0.0	0.0	0	0
1	1	PAYMENT	1864.28	C1666544295	21249.0	19384.72	M2044282225	0.0	0.0	0	0
2	1	TRANSFER	181.00	C1305486145	181.0	0.00	C553264065	0.0	0.0	1	0
3	1	CASH_OUT	181.00	C840083671	181.0	0.00	C38997010	21182.0	0.0	1	0
4	1	PAYMENT	11668.14	C2048537720	41554.0	29885.86	M1230701703	0.0	0.0	0	0

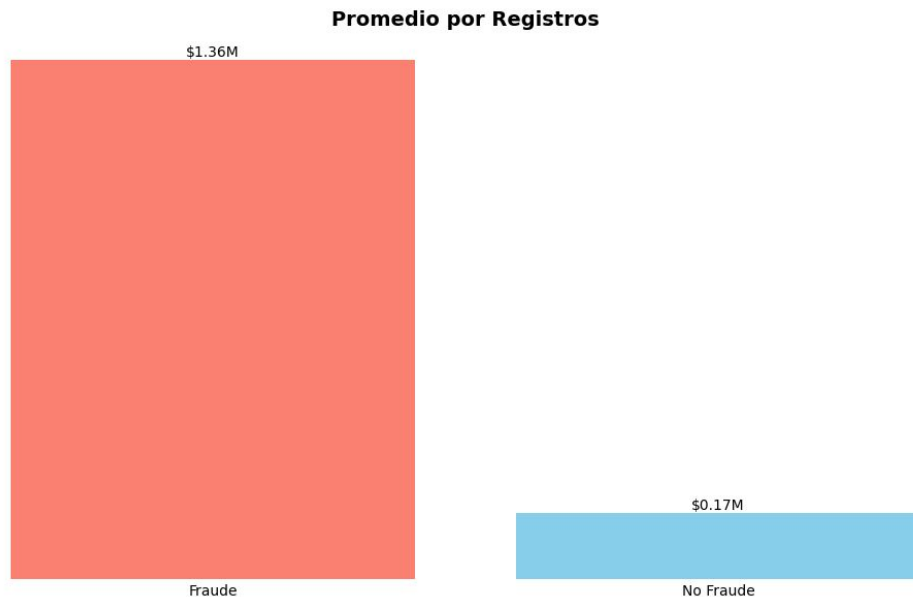


	amount	isFraud	type_CASH_IN	type_CASH_OUT	type_DEBIT	type_PAYMENT	type_TRANSFER	type2_CC	type2_CM	day	hour
0	9839.64	0	0	0	0	1	0	0	1	1	2
1	1864.28	0	0	0	0	1	0	0	1	1	2
2	181.00	1	0	0	0	0	1	1	0	1	2
3	181.00	1	0	1	0	0	0	1	0	1	2
4	11668.14	0	0	0	0	1	0	0	1	1	2



Exploración de Datos

De los registros que son fraudulentos, el promedio es de \$1.36M, lo que es aproximadamente 8 veces más que el promedio de los registros no fraudulentos.

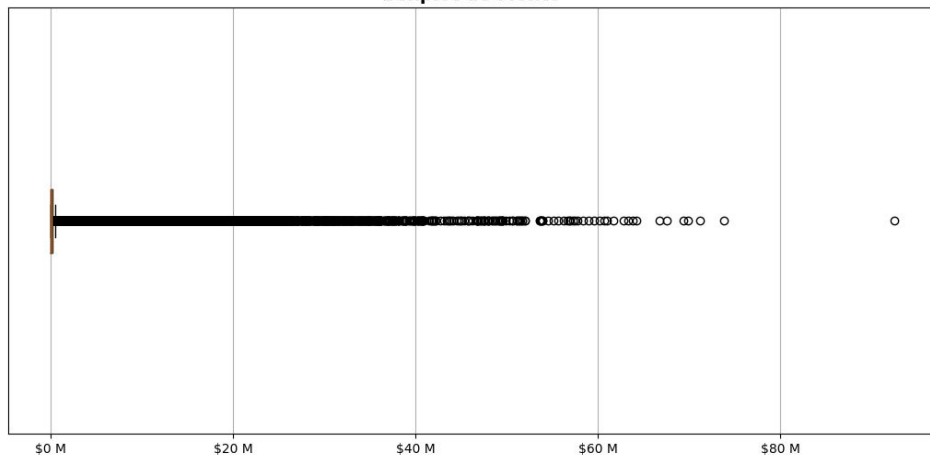




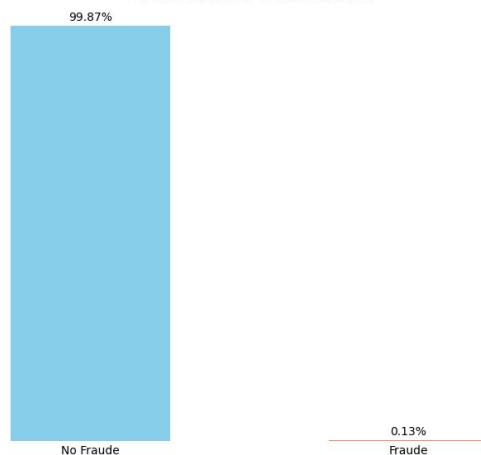
Exploración de Datos

Desbalanceo de clases y Outliers en los montos

Boxplot de Monto



Distribución de Fraudes

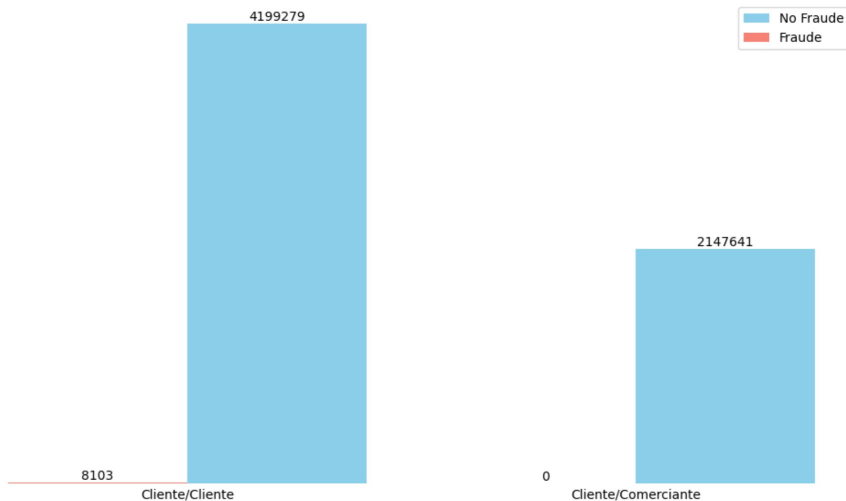




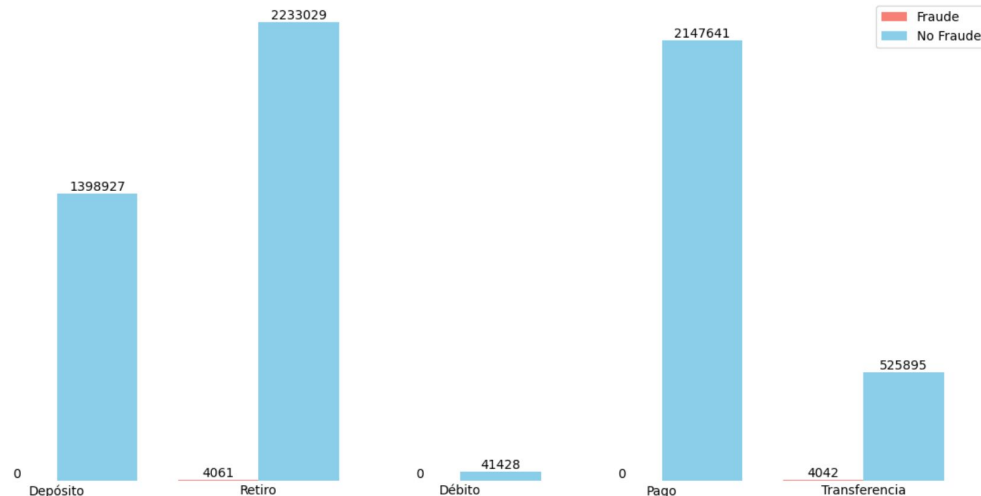
Exploración de Datos

No hay transacciones fraudulentas por **Ciente a Comercio**, ni por los conceptos de: **Depósito, Débito, Pago**.

No. de Transacciones por Destinatario



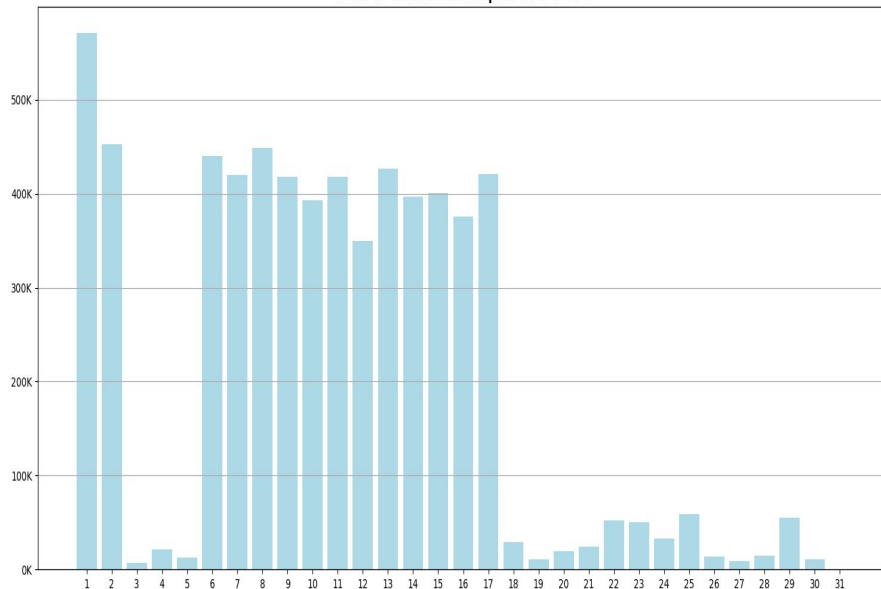
Recuento de Transacciones por Concepto



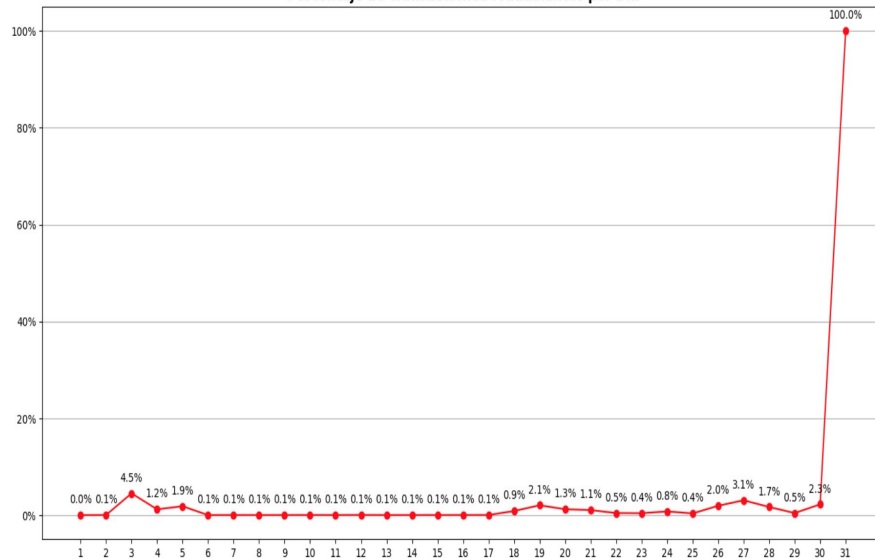


Exploración de Datos

Total de Transacciones por Día del Mes



Porcentaje de Transacciones Fraudulentas por Día

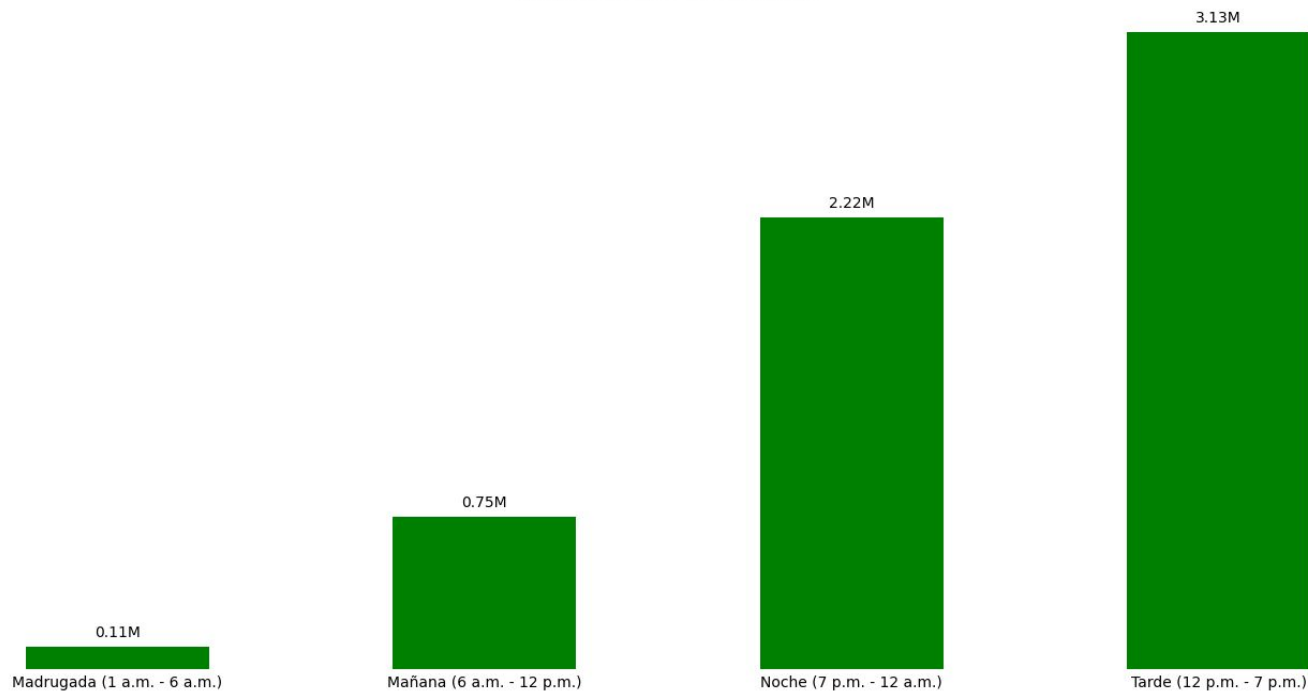


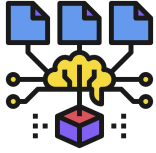
268 fraudes el último día del mes



Exploración de Datos

Transacciones por Horario

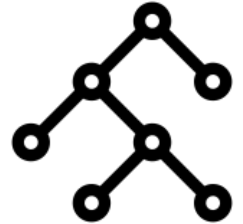
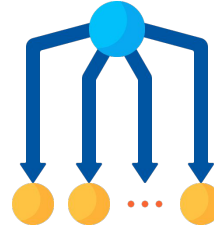
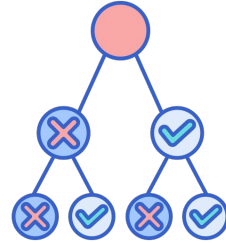
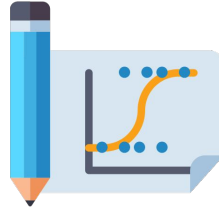


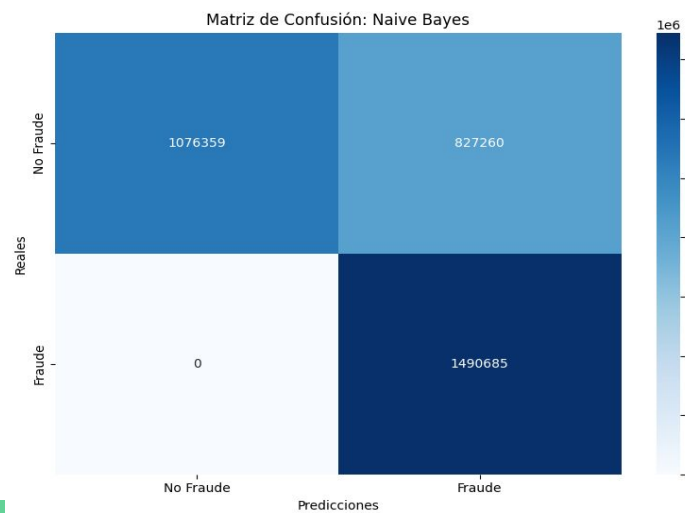
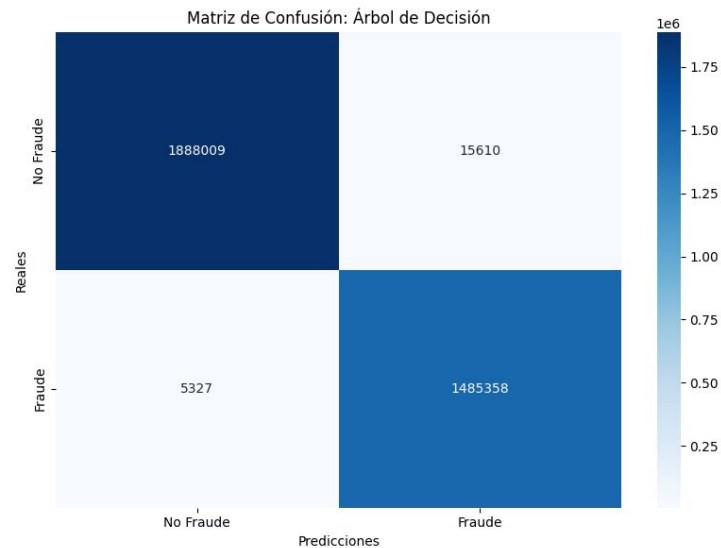
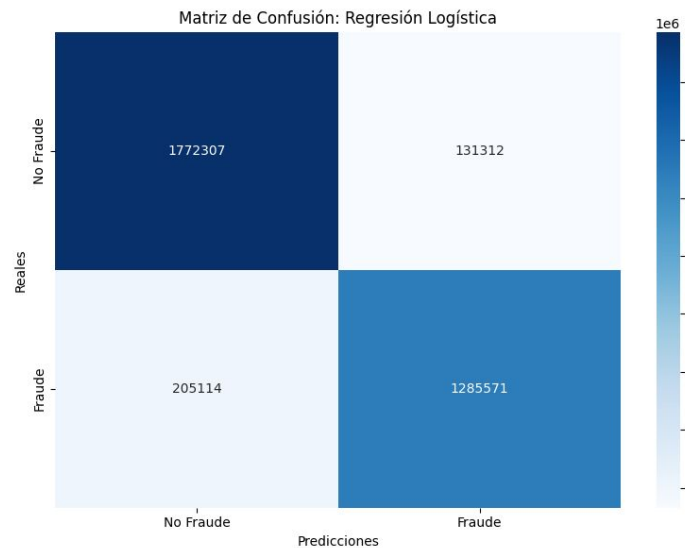


Construcción de Modelos

En este paso se experimentó con 4 algoritmos de machine learning para el caso estudiado.

- Regresión Logística
- Árbol de decisión
- Naive Bayes
- Random Forest

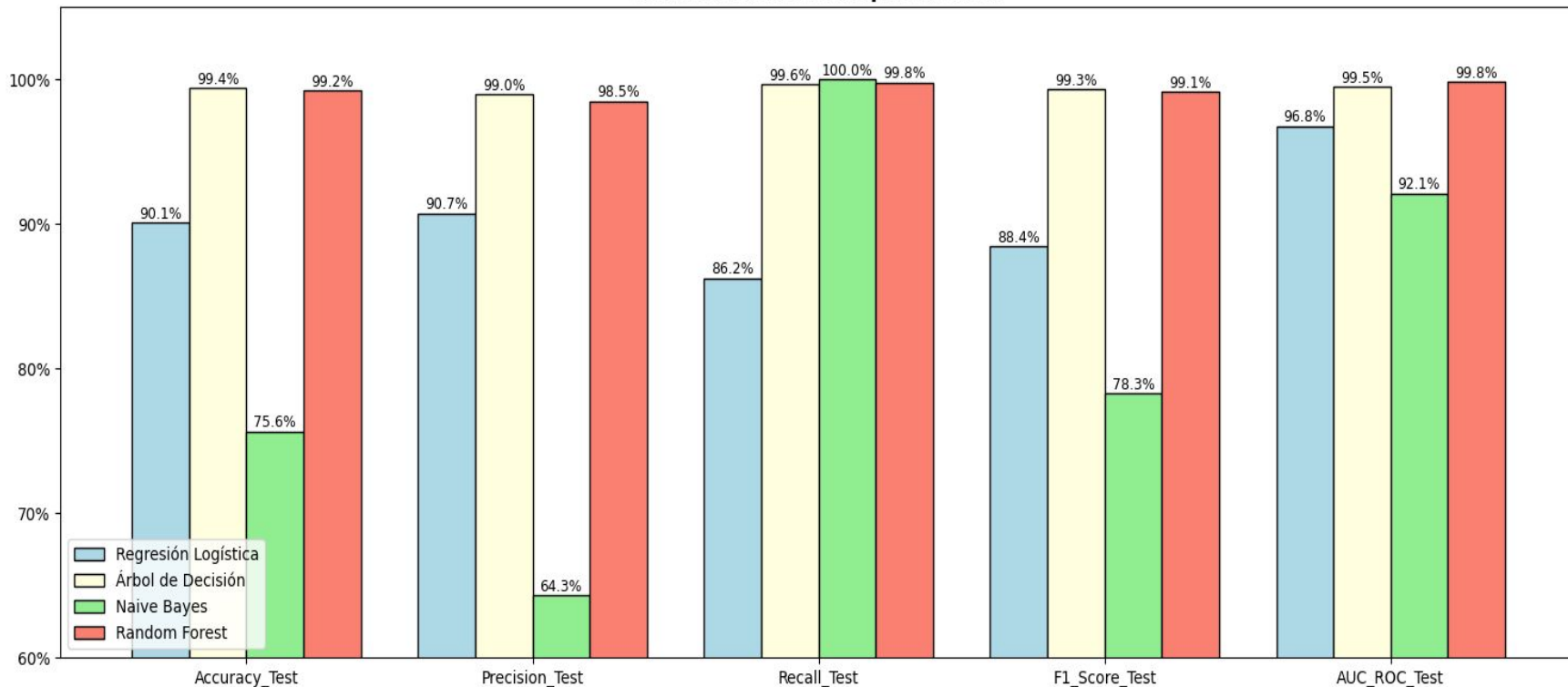






Evaluación del Modelo

Métricas de Prueba por Modelo





Conclusiones

Determinamos que el modelo **Random Forest** era la opción más apropiada debido a sus ventajas específicas en términos de métricas de desempeño y capacidad de generalización en conjuntos de datos grandes, complejos y desequilibrados.

La implementación de este modelo en un entorno bancario puede **mejorar significativamente** la capacidad de detectar y prevenir fraudes, **reduciendo** así las **pérdidas económicas** y **aumentando** la confianza de los **clientes** en la seguridad de las transacciones.

