

Detección de Fraudes en Transacciones Móviles

La detección temprana de transacciones fraudulentas es fundamental para proteger a los usuarios y mantener la confianza en los servicios financieros móviles. Utilizando técnicas avanzadas de machine learning, desarrollaremos un modelo robusto que pueda identificar patrones sospechosos y alertar sobre posibles fraudes.



Preprocesamiento de Datos



Limpieza de Datos

Identificar y manejar valores faltantes, eliminar outliers y corregir inconsistencias en los datos.



Codificación

Convertir variables categóricas en un formato numérico adecuado para el análisis.



Normalización

Escalar y normalizar las variables numéricas para asegurar una adecuada comparación entre ellas.



Selección de Características

Identificar las variables más relevantes para la predicción de transacciones fraudulentas.

Exploración de Datos

1

Análisis Univariado

Examinar la distribución y estadísticas de cada variable por separado para entender su comportamiento.

2

Análisis Bivariado

Estudiar las relaciones entre pares de variables y cómo se influyen mutuamente.

3

Selección de Variables Clave

Identificar las características más relevantes para la detección de fraudes en base a la exploración realizada.



Construcción de Modelos



Regresión Logística

Un modelo de clasificación robusto y bien interpretable, adecuado para detectar fraudes binarios.



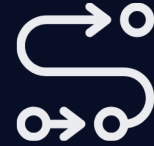
Árboles de Decisión

Capturan fácilmente patrones complejos en los datos y son fáciles de explicar.



Random Forest

Ensemble de árboles de decisión que mejora la precisión y robustez del modelo.



Naive Bayes

Modelo basado en el análisis de probabilidad tanto, con la formulación de Bernouille, Gauss o multinomial.

Evaluación y Selección

1 Accuracy

Mide la proporción de predicciones correctas realizadas por el modelo con respecto al número total de predicciones

2 Precisión

La precisión mide la proporción de instancias correctamente clasificadas como positivas frente al total de instancias que fueron clasificadas como positivas.

3 Recall

La precisión mide la proporción de instancias correctamente clasificadas como positivas frente al total de instancias que fueron clasificadas como positivas.

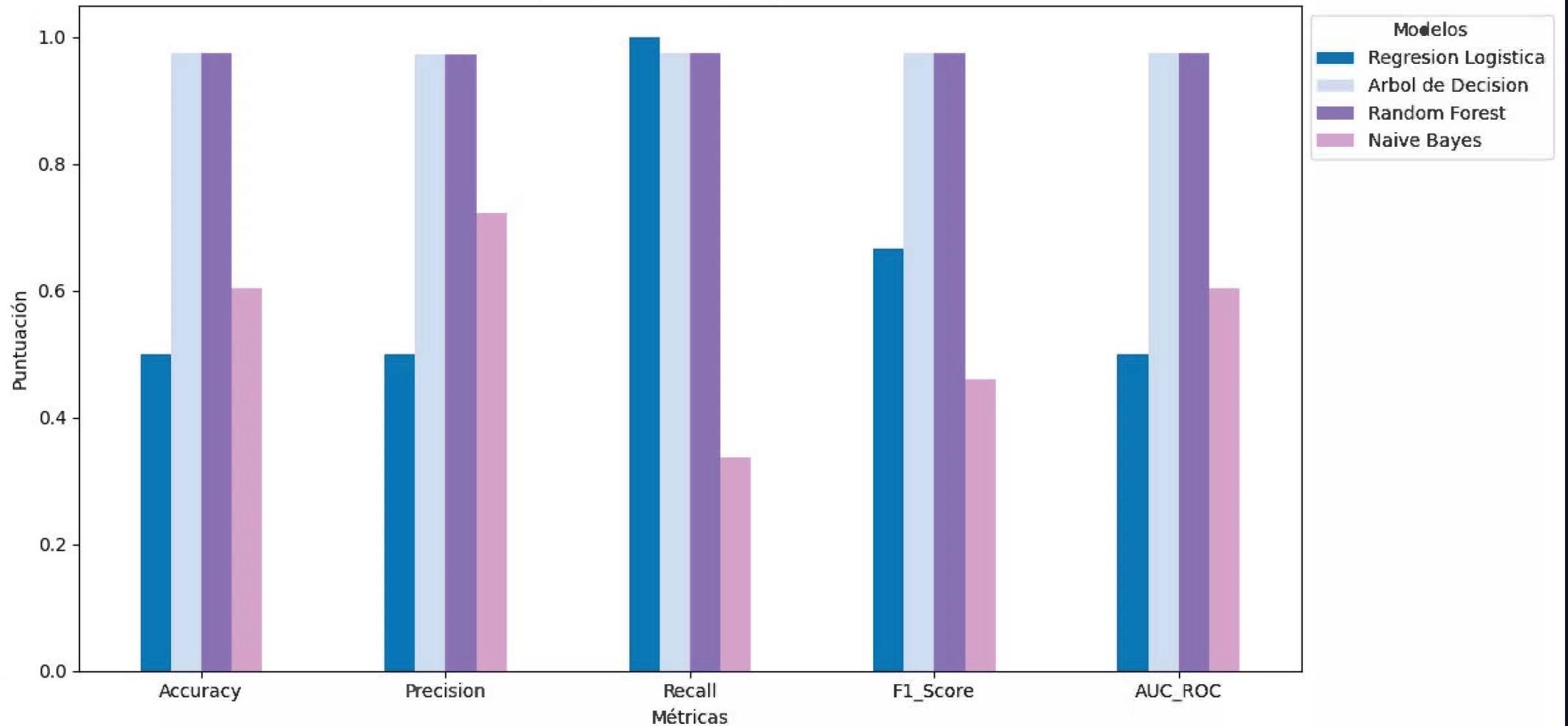
4 Área bajo la curva ROC

La curva ROC traza la tasa de verdaderos positivos (True Positive Rate, TPR) contra la tasa de falsos positivos (False Positive Rate, FPR) para diversos puntos de corte.

5 F1-score

Combina precisión y recall en una sola métrica, dando una visión general del rendimiento.

Comparación de Métricas por Modelo





Conclusión

El modelo de Random Forest se perfila como la opción más apropiada, ya que logra un balance óptimo entre precisión, recall y F1-score en la detección de transacciones fraudulentas. Su capacidad para capturar patrones complejos y adaptarse a la evolución de los datos lo convierte en una solución robusta y escalable para el sector Fintech.