



ESCUELA DE NEGOCIOS

MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA DE DATOS

**APLICACIÓN DE UN MODELO DE DESCUBRIMIENTO DE
ANOMALIAS PARA LA DETECCIÓN DE OPERACIONES
FRAUDULENTAS CON TARJETAS DE CRÉDITO**

BRYAN ANDRES SANAGUANO MEJIA

2023

Introducción

- Desafío crítico en el ámbito financiero
- Implementación de un modelo de descubrimiento de anomalías
- Detectar patrones inusuales de transacciones



Revisión Literaria

Operaciones Fraudulentas



Modelos Estadísticos



Machine Learning



Descripción del problema

La detección de operaciones fraudulentas representa un desafío constante



Incapacidad de los métodos tradicionales de detección de fraudes



Objetivos del Estudio

Objetivo general

- Identificar operaciones fraudulentas en las transacciones con tarjetas de crédito mediante la aplicación de un modelo de descubrimiento de anomalías utilizando técnicas de Machine Learning para proporcionar una herramienta efectiva para la detección temprana y precisa de fraudes en el sector bancario.

Objetivos específicos

- Recopilar y analizar datos históricos de transacciones con tarjetas de crédito, incluyendo información sobre transacciones legítimas y fraudulentas.
- Desarrollar un modelo para la identificación de anomalías en las transacciones de tarjetas de crédito mediante técnicas de análisis de datos como apoyo al análisis de posibles fraudes.
- Evaluar la eficacia del modelo en la identificación de anomalías en las transacciones de tarjetas de crédito como herramienta de apoyo al análisis de posibles fraudes.

Metodología

Recopilación de Datos



Preparación de Datos



Identificación de variables



Análisis de datos

Análisis descriptivo

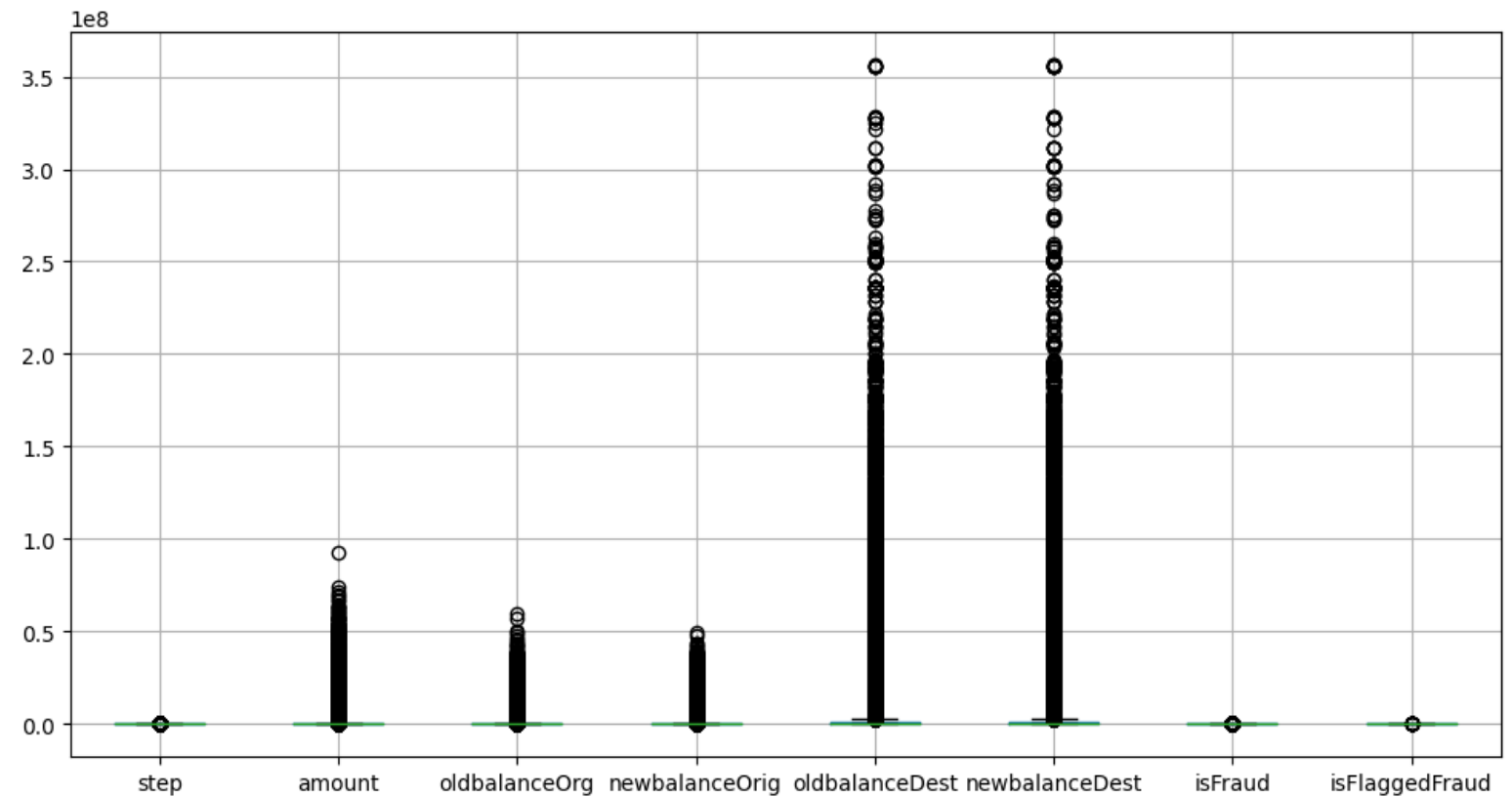
- Resumen estadístico

dtypes	missing#	missing%	uniques	count	min	max	mean	
step	int64	0	0.000000	743	6362620	1.000.000	743.000.000	243.397.246
type	object	0	0.000000	5	6362620	nan	nan	nan
amount	float64	0	0.000000	5316900	6362620	0.000000	92.445.516.640.000	179.861.903.549
nameOrig	object	0	0.000000	6353307	6362620	nan	nan	nan
oldbalanceOrig	float64	0	0.000000	1845844	6362620	0.000000	59.585.040.370.000	833.883.104.074
newbalanceOrig	float64	0	0.000000	2682586	6362620	0.000000	49.585.040.370.000	855.113.668.579
nameDest	object	0	0.000000	2722362	6362620	nan	nan	nan
oldbalanceDest	float64	0	0.000000	3614697	6362620	0.000000	356.015.889.350.000	1.100.701.666.520
newbalanceDest	float64	0	0.000000	3555499	6362620	0.000000	356.179.278.920.000	1.224.996.398.202
isFraud	int64	0	0.000000	2	6362620	0.000000	1.000.000	0.001291
isFlaggedFraud	int64	0	0.000000	2	6362620	0.000000	1.000.000	0.000003

Análisis de datos

Análisis univariado

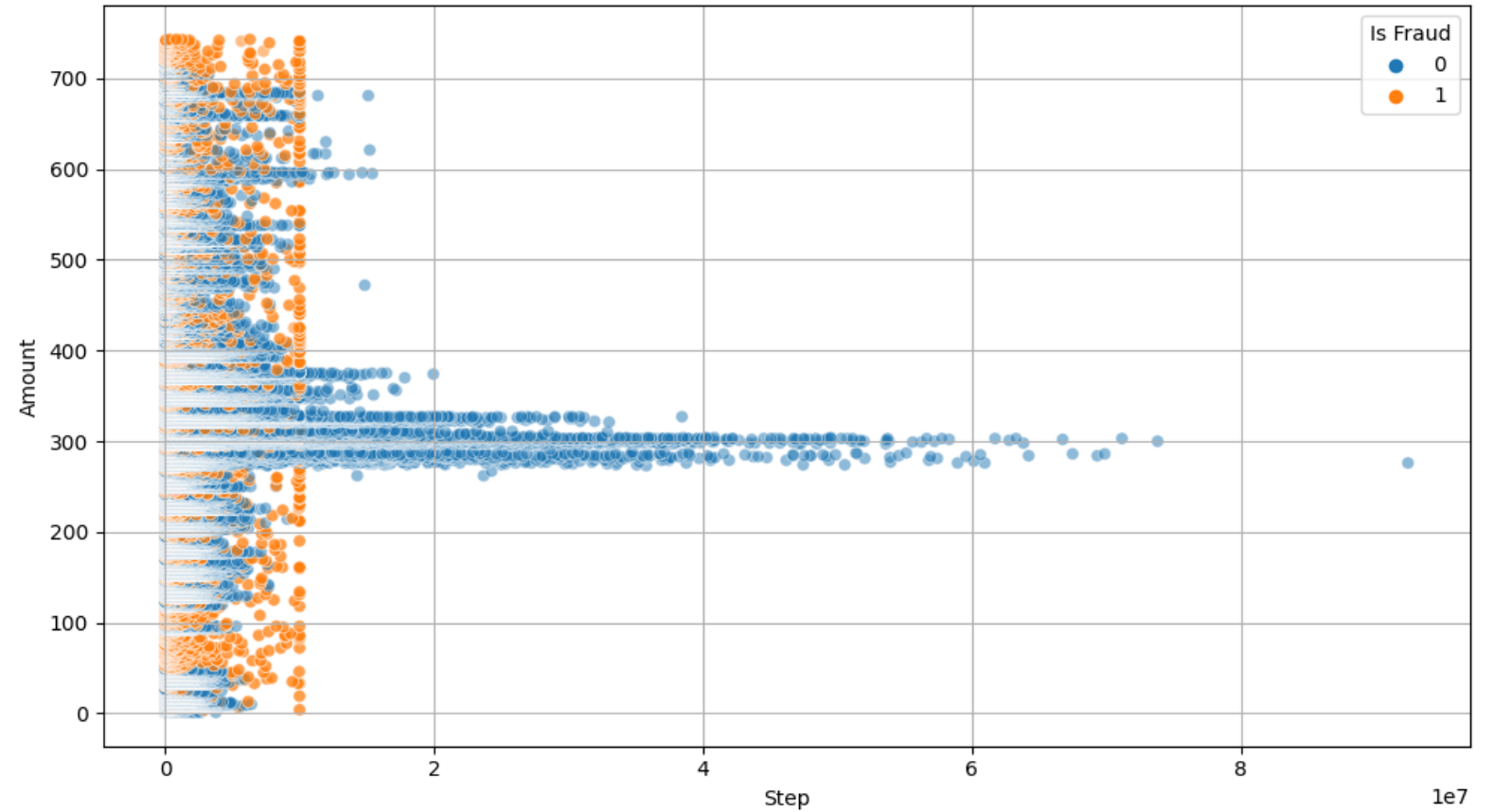
- Booxplot



Análisis de datos

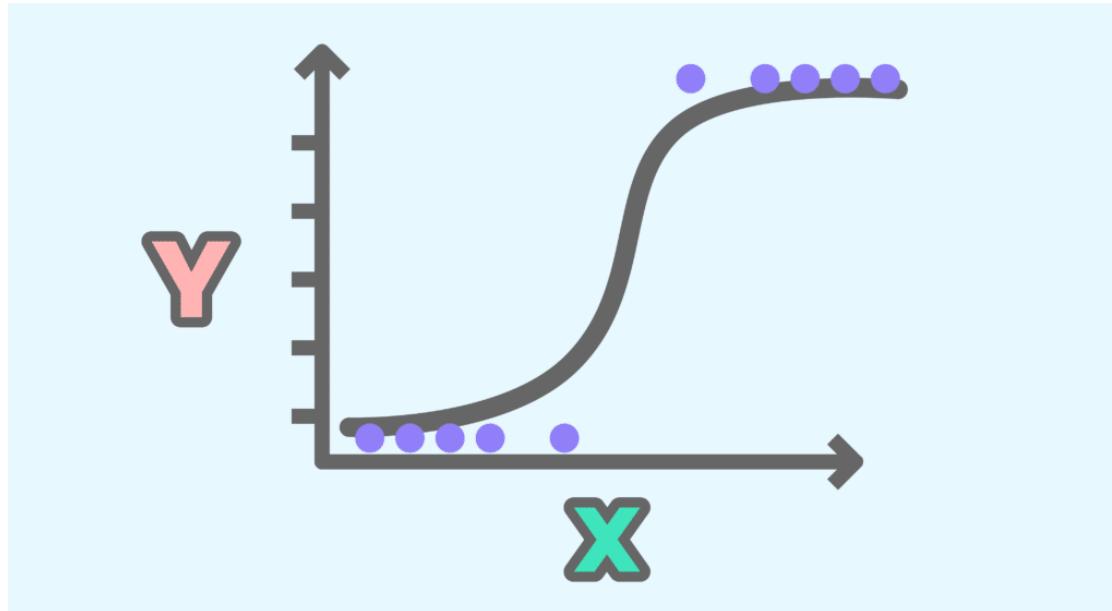
Análisis bivariado

- Dispersión



Descripción del modelo

Modelo de regresión logística



- Modelo estadístico ampliamente utilizado en la detección de fraudes.
- Capacidad para modelar la probabilidad de un evento binario.
- Modela la probabilidad de fraude en función de las variables independientes

Descripción del modelo

Selección de variables



Entrenamiento de modelo



Resultados

Modelo de regresión logística

Métrica	Valor
Exactitud	84.27%
Precision (Clase 0)	90%
Recall (Clase 0)	77%
F1-score (Clase 0)	83%
Precision (Clase 1)	80%
Recall (Clase 1)	92%
F1-score (Clase 1)	85%

- Clase 0: transacciones normales
- Clase 1: transacciones fraudulentas

Matriz de características

	Feature	Coefficient	Absolute_Coefficient
7	type_CASH_OUT	7.292942	7.292942
10	type_TRANSFER	7.076855	7.076855
3	newbalanceOrig	-4.705803	4.705803
1	amount	3.392151	3.392151
9	type_PAYMENT	-2.881433	2.881433
5	newbalanceDest	-0.933098	0.933098
8	type_DEBIT	-0.228908	0.228908
4	nameDest	0.228744	0.228744
0	step	-0.184244	0.184244
6	type_CASH_IN	-0.131955	0.131955
2	nameOrig	0.009294	0.009294

Las características se han ordenado en función de sus coeficientes, lo que indica la fuerza y la dirección de la relación entre cada característica y la variable objetivo

Conclusiones y Recomendaciones

Conclusiones

- La implementación de modelos avanzados, resultó en una mejora significativa en la detección de operaciones fraudulentas con tarjetas de crédito.
- La aplicación de métricas de evaluación rigurosas, como la precisión, el recall y el F1-score, proporcionó una evaluación exhaustiva del desempeño de los modelos.
- La información generada a partir del análisis de datos es valiosa no solo para la detección de fraudes, sino también para la toma de decisiones gerenciales.
- La implementación de modelos de analítica de datos eficaces no solo contribuye a la seguridad de la organización, sino que también puede promover ventajas competitivas.

Recomendaciones

- Los modelos de detección de fraudes deben someterse a un mantenimiento y actualización continuos.
- La detección de fraudes debe ser en tiempo real siempre que sea posible, permitiendo tomar medidas inmediatas para prevenir transacciones fraudulentas y minimizar el impacto.
- Considerar la implementación de múltiples modelos y técnicas de detección de fraudes, al diversificar los enfoques se puede aumentar la robustez y la eficacia de la detección.
- Los modelos de descubrimiento de anomalías no se limitan solo a la detección de fraudes con tarjetas de crédito, se puede considerar la posibilidad de aplicar técnicas similares a otros sectores donde la detección de anomalías sea relevante.