

# La Analítica de Datos y el Comportamiento del Fraude Bancario

Jonathan Estiven Fontalvo Aparicio, Ehider Andres Guarin Ortega, Mario David Diaz Antequera.

Tutor: Wilson Nieto, Ph.D.

Departamento de Ingeniería de Sistemas, Universidad del Norte.

## Introducción

En la era de la globalización digital, los pagos en línea se han vuelto tan comunes como el uso del efectivo, lo cual ha incrementado las preocupaciones sobre delitos financieros.

La prevención y detección de fraude son ahora vitales para la sostenibilidad de los negocios de comercio electrónico. Aprovechando el análisis de grandes volúmenes de datos, las herramientas de Machine Learning y la inteligencia artificial se presentan como soluciones efectivas para estos problemas.

Esta investigación se enfoca en probar diversos modelos de ML para identificar el más preciso y desarrollar un prototipo de servicio para prevenir y detectar fraudes en transacciones online como pagos, transferencias, depósitos, retiros y débitos.

## Descripción del Problema

El fraude en las transacciones en línea es una preocupación creciente en la era digital, afectando tanto a empresas como a usuarios.

Las causas incluyen la falta de seguridad en los sistemas de pago, debilidades en los métodos de autenticación y la ausencia de herramientas efectivas para la detección temprana de fraudes.

Además, muchos usuarios desconocen las tácticas fraudulentas y las señales de alerta, lo que facilita la perpetración del fraude.

## Justificación

Tradicionalmente, la detección de fraude se ha basado en técnicas manuales, que pueden ser ineficientes. Con el avance tecnológico y el aumento de datos, el uso de Machine Learning permite automatizar y mejorar este proceso.

Estudios recientes han demostrado el potencial de estos algoritmos para identificar patrones de fraude. Este estudio compara diversos métodos de ML para encontrar el modelo óptimo, mejorando la detección de fraudes y mitigando riesgos en el sector financiero.

## Objetivos

Diseñar e implementar un sistema inteligente basado en aprendizaje automático para la detección y prevención de fraudes en transacciones en línea, con el fin de proteger la integridad de las operaciones y garantizar la seguridad de los usuarios.

### Objetivos Específicos

- Modelar y diseñar un sistema de aprendizaje automático que permita procesar y analizar datos transaccionales para identificar patrones asociados con actividades fraudulentas en transacciones en línea
- Crear una arquitectura de solución adecuada que permita procesar y analizar datos transaccionales para identificar patrones asociados con actividades fraudulentas en transacciones en línea
- Implementar el prototipo, incluyendo el modelo de aprendizaje automático y sus diferentes técnicas
- Validar el prototipo del evaluando su efectividad en la detección y prevención de fraudes en transacciones en línea.

## Marco teórico

La necesidad de eficiencia ha impulsado el uso de herramientas de Machine Learning para automatizar y agilizar este proceso.

El crecimiento del comercio electrónico ha aumentado el fraude en transacciones en línea, causando pérdidas monetarias globales. Se necesitan técnicas efectivas para detectarlo y mitigar riesgos.

Los sistemas de detección de fraudes utilizan atributos de transacciones para construir clasificadores que distinguen transacciones fraudulentas de legítimas.

Se emplean tanto métodos supervisados como no supervisados para predecir actividades fraudulentas.

Métodos como árboles de decisión y random forests han demostrado resultados prometedores en la detección de fraudes en transacciones en línea.

El desequilibrio en los datos de transacciones pueden afectar la precisión de los sistemas de detección de fraudes. Técnicas como SMOTE abordan esto creando muestras sintéticas de la clase minoritaria.

## Arquitectura del prototipo

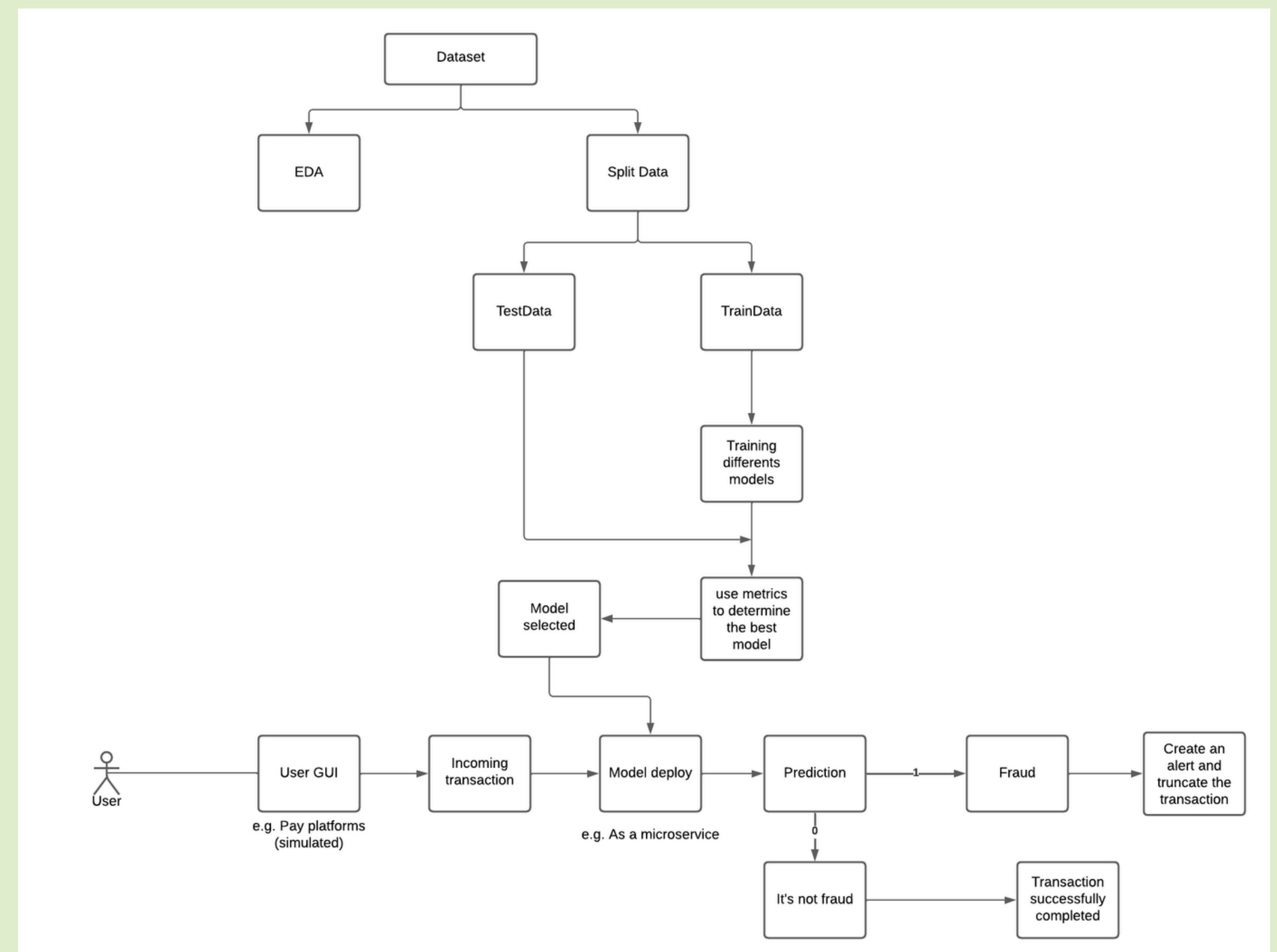


Fig. 1. Arquitectura Lógica de la Solución

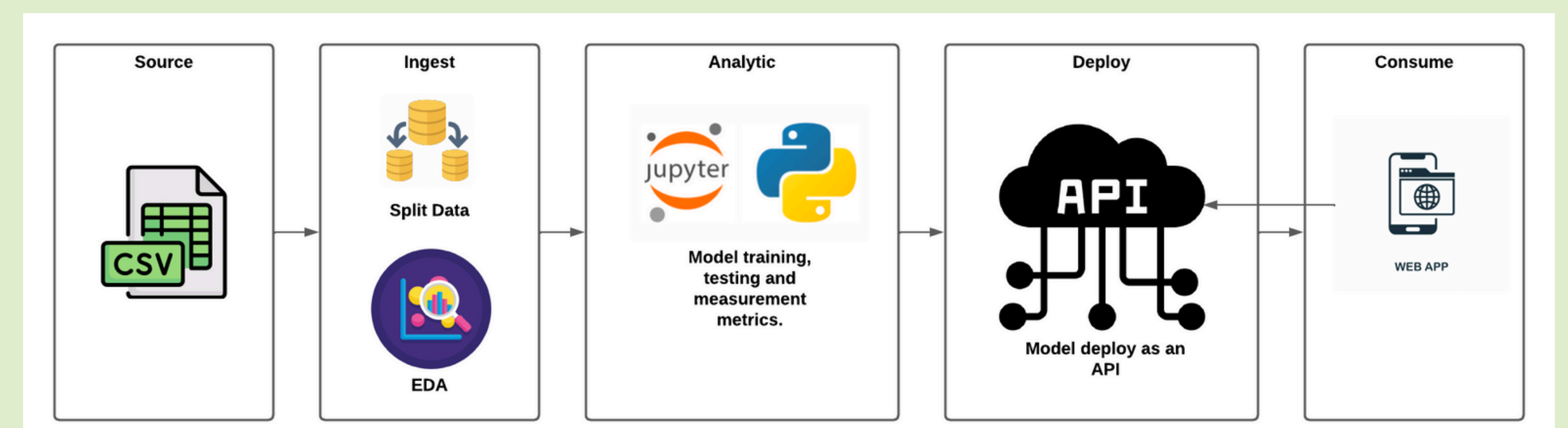


Fig. 2. Arquitectura Física de la Solución

## Prototipo

**Online Payment Fraud Detection using Machine Learning**

Importando librerías

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import drive
import plotly.graph_objects as go
drive.mount('/content/drive')
matplotlib inline
```

Get the Data

Use pandas to read 'data.csv' as a dataframe called df.

```
df = pd.read_csv('PS_20174392719_1491284439457_log.csv')
```

Check out the info(), head(), and describe() methods on data.

Contexto

Existe una falta de conjuntos de datos públicamente disponibles sobre servicios financieros y especialmente en el ámbito emergente de las transacciones de dinero móvil. Los conjuntos de datos financieros son importantes para muchos investigadores y en particular para nosotros que realizamos investigaciones en el ámbito de la detección de fraudes. Parte del problema es la naturaleza intrínsecamente privada de las transacciones financieras, lo que lleva a la falta de conjuntos de datos públicamente disponibles.

Escanear el QR para ver el prototipo completo en GitHub



## Resultados

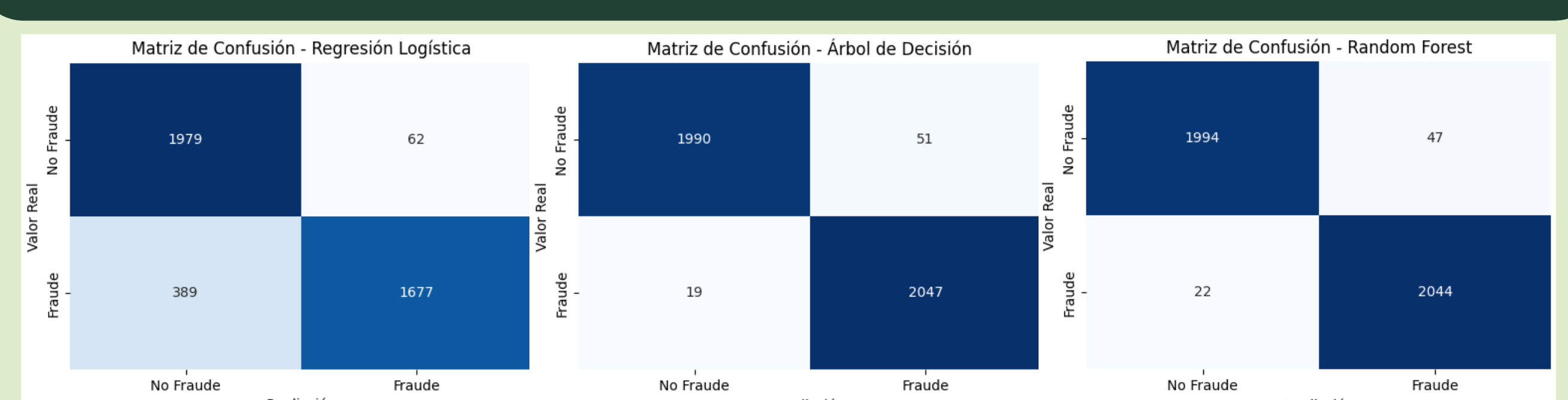


Fig. 3. Métricas obtenidas utilizando Random Undersampling

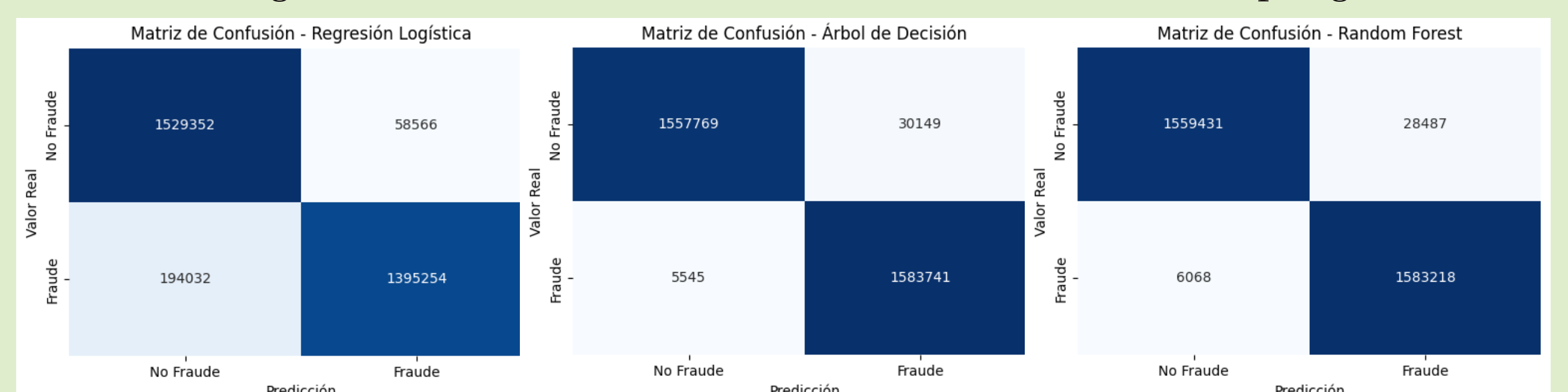


Fig. 4. Métricas obtenidas utilizando SMOTE

## Conclusiones

- El análisis detallado de los datos y la implementación de técnicas avanzadas como SMOTE han permitido identificar las variables más influyentes en la detección de fraudes, mejorando así la precisión y robustez del modelo. Abordar el desequilibrio de clases y utilizar herramientas de Machine Learning demostró ser crucial para crear un sistema de detección de fraudes eficaz y confiable.
- Los modelos de árboles de decisión y Random Forest mostraron un rendimiento superior, destacándose como métodos idóneos para este propósito. La integración de estas técnicas no solo mejora la capacidad de detección, sino que optimiza la efectividad del sistema, subrayando la viabilidad del uso de Machine Learning en la prevención de fraudes en transacciones en línea.

## Bibliografía

A Cost-sensitive weighted Random Forest Technique for Credit Card Fraud Detection. (2019, 1 de julio). IEEE Conference Publication | IEEE Xplore. <https://ieeexplore.ieee.org/document/8944885>

Abakarim, Y., Limam, M., & Attioui, A. (2018). An Efficient Real Time Model For Credit Card Fraud Detection Based On Deep Learning. ACM. <https://doi.org/10.1145/3289402.3289530>

Abdulalem, A., Shukor, A. R., Hajar, O. S., Elfadil, E. T., Arafat, A. D., Maged, N., Tusneem, E., Hashim, E., & Abdu, S. (2022). Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. Applied Sciences, 12, 9637. <https://doi.org/10.3390/app12199637>

Ali, A., Razak, S. A., Othman, S. H., Eisa, T. A. E., Aldhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. Applied Sciences, 12(19), 9637. <https://doi.org/10.3390/app12199637>

Hu, J., Hu, R., Wang, Z., Li, D., Wu, J., Ren, L., Zang, Y., Huang, Z., & Mei, W. (2023). Collaborative Fraud Detection: How Collaboration Impacts Fraud Detection. In Proceedings of the 2021 1st International Conference on Control and Intelligent Robotics. <https://doi.org/10.1145/3581783.3613780>

Amarasinghe, T., Aponso, A., & Krishnarajah, N. (2018). Critical Analysis of Machine Learning Based Approaches for Fraud Detection in Financial Transactions. In Proceedings of the 2021 1st International Conference on Control and Intelligent Robotics. <https://doi.org/10.1145/3231884.3231894>

Alshameri, F., & Xia, R. (2023). Credit card fraud detection: an evaluation of SMOTE resampling and machine learning model performance. International Journal Of Business Intelligence And Data Mining, 23(1), 1-13. <https://doi.org/10.1504/ijbidm.2023.131791>

Elreedy, D., & Atiya, A. F. (2019). A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance. Information Sciences, 505, 32-64. <https://doi.org/10.1016/j.ins.2019.07.070>

Bibliografía Completa:

