Bootcamp Data Science - Ejercicio S20 - Gerardo Rodríguez

Contexto: Trabajas en un banco y te solicitan desarrollar un modelo de detección de fraudes en transacciones de tarjetas de crédito. Tienes acceso a los datos sobre transacciones históricas de tarjetas de crédito. El objetivo es construir un modelo de aprendizaje automático que pueda identificar transacciones fraudulentas y ayudar a prevenir pérdidas para el banco y sus clientes.

Preguntas:

1. ¿Cuáles son algunas características importantes que podrían estar presentes en las transacciones fraudulentas?
Los datos están desbalanceados por lo que es difícil entrenar modelos sin usar sobremuestreo y submuestreo.

Pueden tener patrones normales o anómalos que podrían ayudar a identificarlos: a) Por transacciones (frecuencia, monto, y secuencia), b) Por comportamiento del usuario (inicio de sesión, navegación, hábitos de consumo), c) Por geografía (ubicaciones inusuales, cambios de ubicación repentinos).

2. ¿Qué tipo de algoritmos de aprendizaje automático podrían ser útiles para abordar este problema?
Algoritmos enfocados en detección de fraudes (Isolation Forest, One-Class SVM, y Redes Neuronales Autoencoder).

Redes Neuronales y Aprendizaje Profundo (Redes neuronales convolucionales (CNN) para análisis de patrones en datos estructurados y redes neuronales recurrentes (RNN) para datos secuenciales).

Herramientas de Inteligencia Artificial y Machine Learning (AI en la Detección de Fraude: Aplicación de algoritmos de IA para automatizar la detección de patrones fraudulentos, Modelos Predictivos y de Clasificación: Utilizados para evaluar riesgos de transacciones y clasificarlas como legítimas o sospechosas.

SMOTE Tomek – ayuda a mejorar la generalización de los datos desbalanceados, eliminando ruido y generando datos sintéticos.

3. ¿Cómo evaluarías la efectividad de tu modelo en la detección de fraudes?

Durante el entrenamiento del modelo, ingresaríamos los datos de prueba (identificados como transacciones fraudulentas) a nuestros modelos para validar que el sistema pueda identificarlos.

En la vida real se requiere estar monitoreando modelos experimentales por diversos métodos no solo para prevenir el fraude cuando a nuestro modelo se le escape algo, sino también para ajustar el mismo modelo hasta que se confirme que es confiable.

4. ¿Cuáles son los desafíos comunes en la prevención de fraudes y cómo podrían abordarse?

La necesidad de identificar fraudes en tiempo real.

La constante actualización de los modelos ya algoritmos para seguirle el paso a los hackers.

Los datos desbalanceados que deben adaptarse de manera artificial para tener representatividad de los fraudes que queremos medir.