Reto NDS
Equipo 31 - TeamCheems

El comercio electrónico ha crecido exponencialmente en los últimos años y la pandemia está ayudando a que más personas prueben este servicio. Es por esto que el fraude en el e-commerce está aumentando y, por lo tanto, la seguridad que el usuario tiene para seguir consumiendo en este tipo de servicio disminuye.

En un intento para detectar este tipo de fraude se están implementando distintas soluciones como la gráfica, modelos de aprendizaje automático, aprendizaje profundo, entre otros. Nuestro equipo considera que la solución a este problema es de valor ya que muchas empresas destinarán parte de sus ingresos a combatir este tipo de ataques mientras el comercio electrónico siga expandiéndose.

Dataset

Utilizaremos un conjunto de datos de transacciones de tarjetas de crédito simuladas que contienen transacciones legítimas y fraudulentas desde el 1 de enero de 2019 hasta el 31 de diciembre de 2020. Cubre las tarjetas de crédito de 1000 clientes que realizan transacciones con un grupo de 800 comerciantes. Estos datos se generaron utilizando una herramienta de GitHub llamada Sparkov Data Generation, creada por Brandon Harris.

Información sobre el simulador: https://github.com/namebrandon/Sparkov_Data_Generation
Fuente del dataset: https://www.kaggle.com/kartik2112/fraud-detection

El dataset se divide en dos archivos:

- fraudTrain.csv: contiene más de un millón de registros para entrenamiento.
- fraudTest.csv: contiene más de quinientos mil registros para pruebas.

Cada archivo contiene las siguientes columnas:

Columna	Descripción
ndex	Identificador único
trans*date*trans_time	Día y tiempo de la transacción
cc_num	Número de la tarjeta de crédito del cliente
merchant	Nombre del vendedor
category	Giro del vendedor
amt	Monto de la transacción

first	Nombre del dueño de la tarjeta
last	Apellido del dueño de la tarjeta
gender	Género del dueño de la tarjeta
street	Calle del dueño de la tarjeta
city	Ciudad del dueño de la tarjeta
state	Estado del dueño de la tarjeta
zip	Código postal del dueño de la tarjeta
lat	Latitud del dueño de la tarjeta
long	Longitud del dueño de la tarjeta
city_pop	Número de habitantes de la ciudad donde vive el dueño de la tarjeta
job	Trabajo del dueño de la tarjeta
dob	Fecha de nacimiento del dueño de la tarjeta
trans_num	Numero de transaccion
unix_time	Cantidad de segundos transcurridos desde la medianoche UTC del 1 de enero de 1970
merch_lat	Latitud del comercio
merch_long	Longitud del comercio
is_fraud	Determina si la transacción fue fraudulenta

Para que cualquiera pueda tener entendimiento de estos datos se pretende hacer un dashboard en el que será fácil identificar qué características son importantes al momento de considerar una transacción como una anomalía.

Metodología

Para el modelo de Ecommerce, una herramienta muy poderosa son las reglas de asociación, dado que son consideradas incluso más eficientes en cuanto exactitud, la desventaja es el tiempo de entrenamiento según el algoritmo. No obstante para garantizar la privacidad de los datos de cada usuario y disminuir el tiempo

Algoritmo de CPAR

El algoritmo de CPAR permite concebir un método de clasificación basado en asociación que aportará mayor eficiencia en el tiempo necesario para procesar conjuntos de datos de grandes proporciones. Este algoritmo se basa en el algoritmo FOIL para generar un conjunto menor, pero más efectivo, de reglas.

Algoritmo de MMAC

Este algoritmo trata la existencia de múltiples etiquetas en la clase de problemas de clasificación, en un contexto de asociación, por lo que la ventaja es que se pueden crear reglas de asociación con un modelo de clasificación con clases múltiples, aunque el tiempo de ejecución es bastante extenso siendo un algoritmo de búsqueda exhaustiva.

Algoritmo de FURIA

El algoritmo de Fuzzy Unordered Rule Induction Algorithm, o Algoritmo de Inducción de Reglas de Asociación Borrosas no Ordenadas. Este algoritmo es muy útil para crear reglas de asociación con el enfoque de lógica borrosa, incluyendo un clasificador de las reglas de asociación, lo que será muy útil para comparar el rendimiento de los modelos con el supuesto de que no existe ningún conjunto borroso en nuestro conjunto de reglas.

Entregable

La propuesta consiste en dos evaluaciones de seguridad. Primeramente, con la ayuda de un modelo de reglas de asociación se evaluaría una compra en una página de e-commerce de forma local. Se determinará si el comportamiento del cliente es anómalo, de ser así se rechazaría la compra. Para esta parte los modelos propuestos serán consumidos haciendo uso de una API en python usando la librería FastAPI. En caso que la primera evaluación lo considere un comportamiento normal se ejecutaría la evaluación bancaria. Esto supondría dos evaluaciones aumentando la seguridad del cliente.

Web App-Simulación de página de e-commerce.

Se desarrollará una aplicación web que tenga la función de un ambiente interactivo de prueba. Esta simulará una página de e-commerce donde el usuario podrá realizar compras, y con base en el modelo de reglas de asociación propuesto se identificará como una anomalía o no. En el caso de que se clasifique como una compra anómala, se etiquetara como fraudulenta y se terminaría el proceso.

Modelo Financiero

En el caso en el que no sea etiquetada como fraudulenta, se iniciará la segunda etapa del sistema de seguridad. En este se evalúan los datos bancarios y financieros del cliente en materia de transacciones digitales. Esta segunda etapa se basa en el uso de un algoritmo de machine learning de detección de anomalías como BRMiner, con el cual se realizará una última predicción más rigurosa para asegurar protección al cliente.

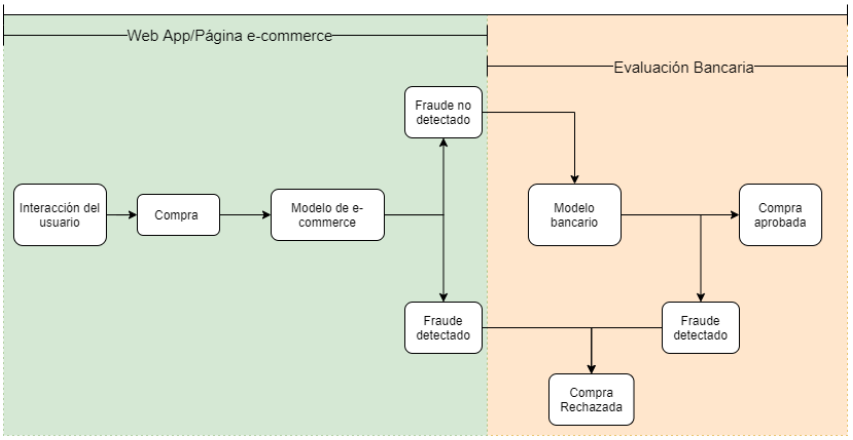

Diagrama del sistema de seguridad propuesto.

Limitaciones y mejoras

La primera limitación muy clara es no poder contar con información específica de nuestro país. Pero entendiendo que nos encontramos en un mundo globalizado, este tipo de fraude se replicará en otras partes del mundo. Nuestra intención es establecer una arquitectura que pueda ser aplicada por sitios de e-commerce para protegerse contra el fraude.