Machine Learning

Caso práctico

Gerard Díaz Hoyos

KeepCoding
Bootcamp de Ciberseguridad
5a edición

El el presente documento se pretende realizar un informe descriptivo de caso de uso de *Machine Learning* en una situación real, con una problemática existente, analizando aquellos puntos a tener en cuenta para la implementación de una solución tecnológica de esta índole.

Se pretende otorgar una visión general y previa a la implantación técnica y puesta en producción.

Detección de fraude con tarjetas de crédito/débito

Descripción caso de uso

1) Contexto general de la problemática:

El fraude en los pagos con tarjeta de crédito/débito en cualquier comercio físico es un problema común en el que se realizan transacciones fraudulentas con tarjetas robadas o clonadas.
El uso de este tipo delictivo puede causar pérdidas significativas al comercio, además de dañar su reputación, con las directas consecuencias que esto puede conllevar.

Este problema no sólo afecta a los comercios en sí en los que se lleva a cabo semejante acción, sino también a los clientes, ya que pueden sufrir cargos no autorizados en sus cuentas bancarias y perder dinero, generar intereses por descubierto, obligaciones legales derivadas, etc. No obstante, hay cierto proteccionismo hacia el usuario final, por lo que se establece que si un comercio o empresa ha procesado una transacción fraudulenta con tarjeta de crédito, está legalmente obligado a devolver la suma al titular de la misma. Si el fraude se

cometiera a nivel *online*, sin intervención de un comercio o empresa determinada, el responsable de abonar la cantidad sustraída sería la entidad financiera del cliente.

Estos tipos de estafas pueden producirse por una de estas dos razones:

- Un delincuente se ha hecho con los datos de la tarjeta de crédito de otra persona.
- El titular de la tarjeta podría no estar siendo honesto (fraude amistoso).

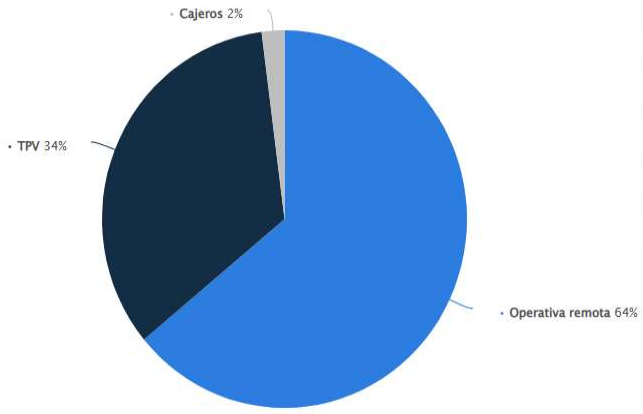
Las formas en las que los defraudadores pueden conseguir los datos de las tarjetas bancarias son variados:

- Robo directo (por ejemplo, los pagos sin contacto no requieren verificación).
- *Skimming* y clonación (acto de hacer copias no autorizadas de los datos de las tarjetas bancarias; con la apropiación de dichos datos, estos podrán utilizarse para realizar compras en línea o duplicados físicos de la tarjeta original; según Nilson Report, estas dos prácticas cuestan a los titulares una media de 28.650 millones de dólares al año).

- Robo de cuentas (que tienen asociadas tarjetas de crédito o débito)
- *Phishing*/Ingeniería Social (comúnmente mediante envíos de comunicaciones de apariencia oficial)
- Infiltración en tiendas *online* legítimas (mediante la inyección de *scripts* en los sitios web de las tiendas *online* atacadas; se realizan mediante herramientas sofisticadas como *MageCart*)

En términos numéricos, podemos exponer las siguientes datos de varias fuentes:

- Según un informe de la Asociación para los Pagos de Europa, en 2020 el fraude en los pagos con tarjeta en Europa ascendió a 1,3 millones de euros.
- El Banco de España recibió un 45% más de reclamaciones en 2020 y la mayor subida se debió a operaciones fraudulentas relacionadas con el uso irregular de tarjetas de crédito y débito (20minutos.es).
- Por cada tres hurtos físicos hay uno de tarjetas de crédito: las pérdidas en comercio por estos delitos suponen 1.800 millones anuales (fuente finales 2021: confilegal.com).
- Según un informa de la compañía de pagos digitales *Stripe*, el costo promedio de un fraude con tarjeta de crédito es de 3,36$ por cada 1$ de transacción, lo que significa que el fraude puede ser muy costoso para los comercios.



Fuente del 2018 (estatista.com)

El problema que se pretende minimizar, por tanto, es el fraude en los pagos con tarjetas bancarias en comercios físicos u *online*. Es importante destacar que la solución del problema no sólo implica la reducción de los costos financieros asociados con el fraude, sino también la mejora de la confianza del cliente y la protección de sus datos personales y financieros. Una solución eficaz puede mejorar la relación entre el comercio y el cliente, aumentar la satisfacción general y, en última instancia, impulsar las ventas y los ingresos.

2) Actualidad de la problemática:

Las medidas de seguridad que se están tomando actualmente son:

- Verificación de la firma del titular de la tarjeta (aunque ésta podría ser falsificada con relativa facilidad)
- Solicitud de identificación del titular de la tarjeta (no tiene que ser suficiente para evitar que un delincuente use una tarjeta robada o clonada)
- Revisión manual de las transacciones sospechosas (puede resultar costosa y no suficiente)

Las tarjetas bancarias también incluyen medidas de seguridad:

- Servicio de Verificación de la Dirección (AVS); servicio diseñado para confirmar la identidad del titular de la tarjeta mirando su dirección registrada.
- *3-D Secure* (3DS); capa de seguridad que pide a los usuarios que introduzcan un código para completar la compra (*Visa Secure*, *SecureCode*, *SafeKey*, etc. según el operador de tarjetas)
- CVV (Valor de Verificación de la tarjeta); diseñador para verificar que la tarjeta está efectivamente en posesión del cliente en el momento de la compra.

Se deberá tener en cuenta que todas estas medidas no siempre resultan efectivas y pueden añadir un cierto nivel de fricción, es decir, que se genere cierta resistencia o inconvenientes en la forma en la que los usuarios

interactúan con el sistema o en cómo se llevan a cabo las operaciones.

Matriz de Confusión:

VP **70%**	**FN** **30%**
FP **2%**	**VN** **98%**

Verdaderos Positivos (VP): casos en los que el sistema detecta correctamente un fraude. Aquí es común encontrar unas tasas de alrededor del 60 – 70% en la detección de fraude en pagos con tarjetas bancarias.
Falsos Positivos (FP): casos en los que indica que hay fraude, aunque en realidad no lo haya. Si hubiera muchos, estaríamos ante un sistema ineficiente y molesto para los clientes mal clasificados. Es común encontrar una tasa de falsos positivos muy baja (2 – 3%).
Verdaderos Negativos (VN): el sistema indica que una transacción es legítima correctamente.
Falsos Negativos (FN): casos en los que el sistema indica que una transacción es legítima, pero en realidad es un fraude (mide la eficacia).

3) Solución:

Creación e implementación de un software especialmente enfocado a la detección de fraudes que pudiera usarse de forma conjunta para las compras físicas en establecimiento y a través del portal web correspondiente, utilizando *machine learning* para realizar predicciones extremadamente rápidas.
El software estaría perfectamente integrado con el proceso de pago mediante TPV o a través de la pasarela de pago *online*, accediendo al sistema de detección contratado y permitiendo o bloqueando el pago, según el resultado de la predicción obtenida. Además, se iría nutriendo una base de datos que los iría guardando para realimentar el sistema e ir contrastándolos con predicciones anteriores.

Por otro lado y, de forma totalmente secundaria, se incluiría un pequeño modelo predictivo especialmente creado para los comercios físicos, basado en un puntaje de riesgo más simple, en el que se podría definir el nivel global de capacidad de detección de fraudes por el comercio en cuestión. Esta solución sólo pretendería ser una ayuda para los encargados de tienda/comercio que permitiera saber, por ejemplo en una escala del 0 al 10, el nivel de probabilidad para detectar un caso de pago fraudulento (tendría en cuenta aspectos tales como entrenamiento del personal, uso de tecnología de chip EMV, monitorización de las transacciones efectuadas, actualización de las políticas de seguridad, etc.)

4) KPIs; indicadores de negocio:

Se puede llegar a medir la efectividad del modelo de detección de pagos fraudulentos con los siguientes indicadores:

- **Tasa de detección de fraudes**: proporción de transacciones fraudulentas que son detectadas por el sistema en relación al total de las procesadas. Un alto porcentaje indicaría eficacia.
- **Tasa de falsos positivos**: un alto porcentaje afectaría a la satisfacción del cliente e incrementaría los costos de operación del negocio.
- **Tasa de rechazo**: proporción de transacciones que son rechazadas por el sistema. Si el porcentaje fuera demasiado alto significaría que el análisis está siendo demasiado restrictivo.
- **Tiempo de respuesta**: un tiempo de respuesta corto será muy importante para evitar pérdidas y minimizar el impacto del fraude en el negocio. El software integrado deberá proveer del resultado de la predicción en el momento en el que se está realizando el pago.
- **Coste del fraude**: indicador de medición del costo total de los fraudes que son detectados.
- **Impacto en la satisfacción del cliente**: un equilibrio perfecto entre la restricción necesaria y la flexibilidad deseada, según los resultados, generará mejor reputación del negocio por parte de los clientes.

5) Mínimos esperados:

Se espera que el modelo de detección de pagos fraudulentos logre reducir la tasa de fraude en los pagos con tarjeta en el comercio en un 50% aproximadamente en el primer año, es decir, deberá ser capaz de identificar la mitad de las transacciones fraudulentas que antes no detectaban. La medición en este caso, será extremadamente difícil de calcular si el software ha sido recientemente implementado en un comercio que no contaba con ningún tipo de medida parecida ni se haya nutrido en los últimos años de una base de datos específica para la problemática que se aborda. Se podrá comparar este mínimo con otros más fácilmente cuantificables que se exponen a continuación.

El coste del fraude se debería reducir también en un 50%, medición más sencilla, tal vez, si se han anotado los gastos jurídicos o relativos a las estafas sufridas.

El tiempo de respuesta se medirá en segundos, y se pretende que sea prácticamente imperceptible en el momento en el que se está realizando el pago. Medirá, por tanto, la efectividad de la integración y se podrá comparar fácilmente con los tiempos que se manejaban anteriormente sin usar el software.

El impacto en la satisfacción del cliente deberá aumentar un 0,5 en las métricas que puntúen la seguridad de la transacción de las encuestas que enviemos a los clientes que han realizado alguna compra o devolución.

6) Validación:

Para decidir si la solución es aceptable, se utilizan criterios cuantificables, tales como la tasa de reducción del fraude en un porcentaje determinado, en este caso del 50% en el primer año, y un descenso del coste económico del fraude en un porcentaje similar. También se tendrán en cuenta las tasas de falsos positivos y falsos negativos, ya que el modelo de detección de fraude implementado no puede ser perfecto en ningún caso y siempre habrá predicciones incorrectas. Será importante establecer un equilibrio entre la reducción del fraude y la tasa de falsos positivos y falsos negativos, ya que si se reducen demasiado estos últimos, se podrían perder oportunidades de ventas legítimas o, si se redujera en exceso el fraude, podría significar que se están pasando por alto transacciones fraudulentas.
Además, se deberá considerar cuántos fallos se pueden permitir y, aún así, considerar la solución aceptable. Esto podría depender de los costos de las acciones y los ingresos incrementales esperados al lograr el objetivo. Se podría determinar el punto de inflexión que maximice la ganancia neta del comercio al encontrar el equilibrio entre los costos de la solución y los ingresos extra obtenidos en un año por la reducción del fraude.

7) Experimentación:

Para corroborar el funcionamiento del modelo de detección de fraude se realizarán pruebas rigurosas de validación cruzada utilizando conjuntos de datos de entrenamiento y validación.

Además, también se llevarán a cabo pruebas en tiempo real en un entorno de producción, utilizando transacciones reales para verificar el rendimiento del modelo en un ambiente realista.

Estas pruebas se realizarán con una regularidad de 6 meses, además de lanzar la recomendación de hacerlas también en la mitad de temporada alta (la temporada alta podrá ser diferente dependiendo del producto central del comercio donde se implante este sistema de detección de fraudes en los pagos).

El tiempo necesario para verificar el funcionamiento del modelo dependerá de la cantidad de datos de prueba utilizados y de la complejidad del modelo.

8) Productivización:

Se ofrecerán 3 modos:

- Integración del modelo en el sistema de pago del comercio, de modo que las transacciones se revisen automáticamente en el momento en que se produzcan, y se informe al personal del comercio en caso de detectar una transacción sospechosa. Esta integración puede requerir la colaboración del equipo

IT del comercio o empresa para asegurar que la solución se integre adecuadamente en el sistema existente.

- Entrega de la solución en forma de fichero de texto, que se entregará con una periodicidad previamente pactada para que los propios usuarios lo integren en su propio sistema de detección de fraude. Se deberá comprobar que el fichero sea compatible con los sistemas de los usuarios y se deberá proporcionar instrucciones claras para su uso e implementación.
- Ofrecimiento para desplegar una interfaz gráfica para el usuario en sus propios sistemas informáticos, que permita la visualización de los resultados del modelo en tiempo real y facilite la toma de decisiones en cuanto a la gestión de transacciones sospechosas.

Equipo de trabajo

Podría ser interesante contar con los roles siguientes:

- ***Data Scientist***: encargado de desarrollar el modelo de detección de fraude y llevar a cabo el análisis de datos. Escogerá las variables más relevantes para el modelo y desarrollará o escogerá el algoritmo adecuado.
- ***Business Intelligence***: analizará los datos y la información del negocio del cliente y proporcionará recomendaciones para mejorar el modelo de detección de fraude. Trabajará en estrecha colaboración con el equipo de *Data Science* para asegurarse de que el modelo tenga en cuenta todas las variables relevantes del negocio.
- I**ngeniero de software**: encargado de implementar el modelo de detección de fraude en el sistema de pago del comercio y garantizar su correcto funcionamiento. Será responsable de integrar el modelo con el sistema existente y desarrollar cualquier código adicional necesario.
- **Especialista en ciberseguridad**: detectará las posibles brechas de seguridad existentes, realizando pruebas de penetración y análisis de vulnerabilidades.
- **Responsable de proyecto**: llevará la coordinación de todo el equipo de trabajo y hará un seguimiento de la consecución de los objetivos y plazos preestablecidos. Aunque no será el único, mantendrá la comunicación necesaria con el cliente y le informará de todo aquello relacionado con el progreso del proyecto.
- **Marketing**: funciones típicas tales como definición de la estrategia de comunicación, promoción de la solución, acciones de relaciones públicas, etc. Además, podrían contribuir en la recopilación y análisis de datos de los clientes para mejorar la efectividad del modelo.

Detalle del caso de uso

1) Detalle funcional:

El caso de uso se centrará en detectar y prevenir el fraude de pagos con tarjeta de crédito/débito en un comercio determinado.
El sistema deberá ser capaz de detectar transacciones sospechosas y alertar al equipo de seguridad para que se puedan tomar medidas preventivas.
Se utilizarán algoritmos de *machine learnin*g para analizar los patrones de uso de las tarjetas y detectar transacciones anómalas.
Se llevará un registro de todas las transacciones y los resultados de su análisis, para poder realizar un seguimiento y mejorar el rendimiento del sistema.

Los algoritmos empleados tendrán en cuenta una variedad de parámetros:

- Elementos se seguridad propios de la tarjetas bancarias usadas (AVS, 3DS, CVV, etc.)
- Puntuación de riesgos, con el cual se intentará "cuantificar" o calibrar el riesgo. Al tratarse de un sistema de detección de fraudes en los pagos, se emplearán atajos diseñados para tomar decisiones rápidas utilizando la lógica. Se crea un árbol de decisión combinando varias reglas de riesgo (por ejemplo, usando reglas de velocidad, comparativas entre datos actuales y del histórico...). Cuando la puntuación de riesgo alcanza un determinado umbral, el sistema automatizado decide bloquear o permitir la transacción.
- Enriquecimiento de datos, para confirmar la identidad de un cliente determinado. El enriquecimiento de datos es una capa de seguridad invisible que funciona obteniendo más información de un solo punto de datos (huella digital, análisis IP, búsqueda inversa de redes sociales, etc.). Este proceso se integra perfectamente con el de puntuación de riesgos. Además, ayuda a registrar más información sobre los distintos usuarios.

2) Identificación de orígenes de datos:

Desde un punto de vista funcional, se utilizarán los registros de transacciones de la empresa de tarjetas de crédito/débito, que contendrán información como la fecha, la hora, el importe, el tipo de tarjeta, el número de la tarjeta, el establecimiento en el que se realizó la transacción, y cualquier otro dato relevante. También se utilizarán los registros de ventas del comercio, que contendrán información tal como el producto comprado, el precio, la fecha y la hora de la venta, dirección IP, información sobre el dispositivo utilizado para el pago o para iniciar sesión, dirección de correo electrónico, dirección de facturación, número de teléfono, etc.
Además de estos datos, se pueden incluir otros como el historial de compras del usuario (nos proporciona información sobre los patrones de gasto del usuario; interesante para detectar comportamientos anómalos) y

la ubicación geográfica de la transacción (se puede llegar a considerar sospechoso el hecho de la transacción se haya realizado desde una ubicación inusual).

Desarrollo del caso de uso

1) Seguimiento:

En cuanto a los posibles puntos intermedios o de seguimiento, se pueden realizar diversas pruebas para comprobar la eficacia del sistema de detección de fraudes. Por ejemplo, se puede llevar a cabo un análisis comparativo entre los resultados obtenidos por el sistema y los casos de fraude efectivamente reportados por el personal de seguridad del comercio en el pasado. De esta forma, se podría verificar si el sistema es capaz de detectar los mismos casos que dicho personal o incluso detectar otras casuísticas que no habían sido identificadas anteriormente.

Otra posible forma de realizar un seguimiento efectivo podría ser el análisis de la tasa de falsos positivos y falsos negativos. Es decir, se podría verificar si el sistema está generando alertas innecesarias o si, por el contrario, está dejando pasar transacciones sospechosas. Si se detecta algún problema en este sentido, se podrían hacer ajustes en los algoritmos de detección y predicción para mejorar la precisión del sistema.

Y si encontráramos ciertos patrones no contemplados previamente al investigar los datos que se van recopilando, podríamos incluirlos en el modelo de detección. Por ejemplo, podríamos llegar a observar que las compras en determinadas horas del día son más proclives a ser fraudes o si un número determinado de compras y de cierta cantidad tienen alguna relación también con este tipo de delito.

2) Aporte esperado por Big Data:

El uso de técnicas de *big data* y *machine learning*, permiten una mejora significativa en la detección y prevención de fraudes. Actualmente, estos se basan principalmente en reglas y patrones predefinidos, lo que puede limitar su eficacia y aumentar el número de falsos positivos.

La implementación de algoritmos permite una detección más precisa de patrones y anomalías en las transacciones, lo que reduce el número de falsos positivos y permite una respuesta más rápida ante los casos reales de fraude. Además, la capacidad de análisis de grandes volúmenes de datos permite identificar patrones y tendencias que no podrían detectarse con métodos más tradicionales y observar y controlar posibles sesgos.

También es destacable que el uso de *big data* permite reducir los costos asociados a la gestión y prevención de fraudes, ya que los procesos manuales de revisión de transacciones fraudulentas son sustituidos por un sistema automatizado y más eficiente.

Por último, el uso de *big data* puede ser beneficioso por la posibilidad de recopilar y analizar grandes cantidades de datos para mejorar la comprensión de los patrones de uso de las tarjetas y la identificación de nuevas formas de mejorar la seguridad en el pago con tarjetas de crédito/débito en el comercio.