



THE UNIVERSITY OF
CHICAGO

Proyecto Final

Inteligencia Artificial y Data Science para Directivos
University of Chicago

Estudiante: Walter Enrique Moore
T.A. David Barrero

Santander - Febrero 2022



Problema

Las tarjetas de crédito platino experimentan un nivel elevado de actividades fraudulentas en contraposición a los valores esperados.

Objetivo

Detectar una transacción fraudulenta con la mayor precisión posible y menor cantidad de falsos positivos que ocasionen molestias a los clientes, disminuir los daños de reputación o reclamaciones de entidades de control y regulación.

Solución propuesta

Algoritmo predictivo de detección temprana de transacciones fraudulentas que cancele las mismas hasta su verificación mediante Machine Learning con posibilidad de escalar. Modelo escogido "Random Forest".

Insumos solicitados a las áreas involucradas

Datos existentes en la empresa relacionados con el uso de las tarjetas de crédito, clientes, denuncias de fraude, quejas, métricas, auditorias, proyectos anteriores, informes sobre perdidas en los periodos anteriores por casos de fraude o malas clasificaciones (clientes, económicas, de reputación, etc).

Resultados del equipo técnico

Estudio realizado por nuestros expertos sobre la información suministrada, permitió la obtención de un modelo de ML apto para cumplir con las hipótesis elaboradas por el equipo de Inteligencia y Ciencia de Datos mediante la técnica de Random Forest.

Recomendaciones adicionales basadas en estudio de la competencia y el mercado financiero:

- 1) Enfocar la estrategia comunicacional en la prevención, mediante campañas de concienciación.
- 2) Adopción de múltiples capas de seguridad para la validación de operaciones de usuarios. (mensaje de texto, aplicaciones de celular, tarjeta virtual).
- 3) Alertar sobre la “ingeniería social” a los clientes, el phishing y la suplantación de identidad.
- 4) Incentivar a los clientes a mantener actualizada su información de contacto (teléfonos / email) para notificarlos de manera directa cuando ejecutan operaciones y alertarlos con la mayor brevedad posible en caso de detectar una transacción con sospecha de fraude.

La mejor manera de prevenir es educar al cliente.

Mejora adicional propuesta

Implementar un mecanismo de validación de autenticidad automatizado sobre las operaciones sospechosas. Mediante un mensaje de texto o aplicación de dispositivo móvil que solicite elegir una opción al titular si la transacción marcada con fraude es válida y permitir realizar la operación; o rechazar la misma y proceder al bloqueo de la tarjeta. Disminuyendo al máximo posible el error de clasificación en el funcionamiento del sistema anti fraudes.

Análisis cuantitativo del impacto de negocio

Se efectuó un estudio técnico de diferentes modelos supervisados de ML. Los mejores resultados comparándolo con la situación actual se obtuvieron con el algoritmo “Random Forest”.

Informe financiero	Situación Actual		Random Forest (Solución ML)		Potencial de creación de valor del modelo ML (Ganancia)
	Cantidad (Perdidas)	Económico (Perdidas)	Cantidad (Perdidas)	Económico (Perdidas)	
Fraudes no detectados	1.340 (0,13%)	\$4.020.000 USD	100 (0,01%)	\$300.000 USD	\$3.720.000 USD
Falsos positivos	4.030 (0,40%)	\$1.209.000 USD	100 (0,01%)	\$30.000 USD	\$1.179.000 USD
LTV “Tiempo de vida de cada cliente”	\$18.795.000 USD		\$700.000 USD		\$18.095.000 USD
TOTAL Cuantitativo / Directo	5.370 (0,53%)	\$24.024.000 USD	200 (0,02%)	\$1.030.000 USD	\$22.994.000 USD

Cálculo efectuado para 1.000.000 de clientes. Cada fraude no detectado implica una pérdida de \$3.000 USD. Cada operación clasificada como fraude de manera errónea (falso positivo) implica una pérdida de \$300 USD. Por cada cliente perdido por el mal funcionamiento del sistema implica una pérdida de \$3.500 USD (LTV).

LTV = Customer Lifetime Value es la suma de todas las ganancias (ingresos) que genera el cliente sobre la totalidad de la relación con la empresa de tarjetas (tiempo).

Análisis cualitativo del impacto de negocio

Informe financiero	Situación Actual		Random Forest (Solución ML)		Potencial de creación de valor del modelo ML (Ganancia)
	Cantidad (Perdidas)	Económico (Perdidas)	Cantidad (Perdidas)	Económico (Perdidas)	
Fraudes no detectados & falsos positivos & LTV. Cuantitativo / Directo	5.370 (0,53%)	\$24.024.000 USD	200 (0,02%)	\$1.030.000 USD	\$22.994.000 USD
Costo de las malas decisiones. Cualitativo / Indirecto.	Pérdida de 5370 clientes de la cartera y una alta probabilidad de demandas, pérdida de imagen pública, mala publicidad y daño directo a la reputación de la empresa. La clave de su fracaso, es la alta tasa de falsos positivos y la poca capacidad de detección de operaciones fraudulentas.		Pérdida de 200 clientes de la cartera y una baja cantidad de reclamos, reducción del daño a la imagen pública y reputación. La clave del éxito, es su baja tasa de falsos positivos y alta capacidad de detección de operaciones fraudulentas con posibilidad de mejora mediante técnicas no supervisadas de ML.		(-) Pérdida de Clientes. (-) Fraudes no detectados. (-) Falsos positivos. (-) Pérdida de LTV (-) Daño a la imagen (-) Daño a la reputación (-) Demandas y reclamos. (+) Satisfacción de cliente. (+) Mejora en el servicio. (+) Ganancia empresarial. (+) Seguridad y prestigio.

Cálculo efectuado sobre 1.000.000 de clientes.

Objetivos de mejora para el negocio

- Reducir el 92,54% de las transacciones fraudulentas y el 97,52% del número total de falsos positivos mediante la implementación del modelo de ML.
- Disminuir en \$1.179.000 USD las perdidas económicas por detecciones erróneas que causan un falso positivo y \$3.720.000 USD en las perdidas por operaciones fraudulentas no detectadas.
- Incrementar las ganancias empresariales al retener la mayor cantidad posible de clientes por LTV en \$18.095.000 USD.
- Aumentar la retención de clientes en un 96,28% por un correcto funcionamiento del servicio prestado.
- Amortizar los costos de desarrollo e implementación del modelo de ML y la totalidad de las perdidas por errores en el sistema actual de detección de fraudes en el 100% del posible valor económico perdido de continuar operando hasta su puesta en marcha.

Propuestas futuras iteraciones

Realizar un estudio posterior a la implementación del modelo “Random Forest” de modelos ML basados en aprendizaje no supervisado. Con la finalidad de mejorar el actual proceso de etiquetado de las transacciones y encontrar nuevas estrategias de detección y mitigación de fraudes

Solicitaríamos a futuro a nuestro equipo técnico un análisis para técnicas no supervisadas que incluyeran One Class SVM (máquinas vectoriales) y/o Auto Encoders (redes neuronales).