Inteligencia artificial - CT Innovación

Stripe Radar

Luuk Willemsen
Marc Angel Ortiz
Daniel Vilanova González

Facultat d'Informàtica de Barcelona
Universitat Politècnica de Catalunya

Índice

Inteligencia artificial - CT Innovación **1**

Índice **2**

Introducción: Stripe y Stripe Radar **3**

Stripe 3

Fraude con tarjetas de crédito 4

Stripe Radar **4**

Falsos Negativos 4

Falsos Positivos 5

Beneficios 5

Aprendizaje Automático 6

Entrenamiento **8**

Evaluación de un modelo **9**

Terminología 9

Impacto en la empresa **11**

Bibliografía **13**

Introducción: Stripe y Stripe Radar

Stripe

Stripe es una empresa de tecnología de EEUU que opera en más de 25 países y permite tanto a individuales como a empresas aceptar **pagos** a través de internet **con tarjeta de crédito** de forma segura (similar al servicio que ofrece PayPal).

Ofrece un sistema mediante el cual Stripe almacena los datos de la tarjeta de crédito mientras que la empresa en ningún momento tendrá acceso a ellos (únicamente a un identificador). De esta manera la empresa no tendrá que pasar los rigurosos controles necesarios para aceptar pagos y podrá centrarse en vender su producto o servicio.

Uno de los aspectos a tener en cuenta en el comercio online es el **fraude.** Los métodos antiguos de combatir el fraude online se basan en reglas basadas en la observación, que no son muy eficientes, y aquí es donde entra **Stripe Radar**, un sistema basado en el **Aprendizaje automático** para combatir el fraude online.

Fraude con tarjetas de crédito

El fraude con tarjetas de crédito es muy común. La forma más sencilla es cuando el defraudador obteniendo de alguna manera la tarjeta de crédito de otra persona y comprar un objeto por internet para posteriormente venderlo más barato. El defraudado descubrirá un pago no autorizado y reclamará a su banco una devolución. Normalmente el banco opera a favor del cliente y en este caso el vendedor será responsable injustamente del importe defraudado y los costes adicionales que conlleva la devolución.

Además si a un vendedor tiene una tasa de fraudulencia elevada (más del 1% de las transacciones son fraude) las redes como Visa o MasterCard suelen dejar de aceptar pagos a este vendedor.

Stripe Radar

Stripe Radar es un sistema **automático** de detección y prevención de **fraude online.** Viene incluido en cualquier cuenta de Stripe de forma gratuita y durante un período de prueba de un mes ha supuesto un ahorro de 40 millones de dólares en fraude para Watsi, una ONG que ayuda a financiar tratamientos médicos.

En el fraude online podemos distinguir dos tipos de casos perjudiciales:

Falsos Negativos

Un **falso negativo** se produce cuando no se ha detectado el fraude antes de terminar la transacción. Es decir, se completa el pago.

Falsos Positivos

Un **falso positivo** se produce cuando Stripe no permite un pago legítimo, impidiendo al cliente realizar su compra.

Podríamos evitar todos los **falsos negativos** de forma trivial bloqueando todos los pagos, pero sería catastrófico para el negocio. De la misma manera aceptando todos los pagos reducimos a cero la tasa de **falsos positivos**, pero no avanzamos en la prevención de fraude. Stripe Radar intentará encontrar un balance entre los dos de forma que se minimicen los dos factores.

Beneficios

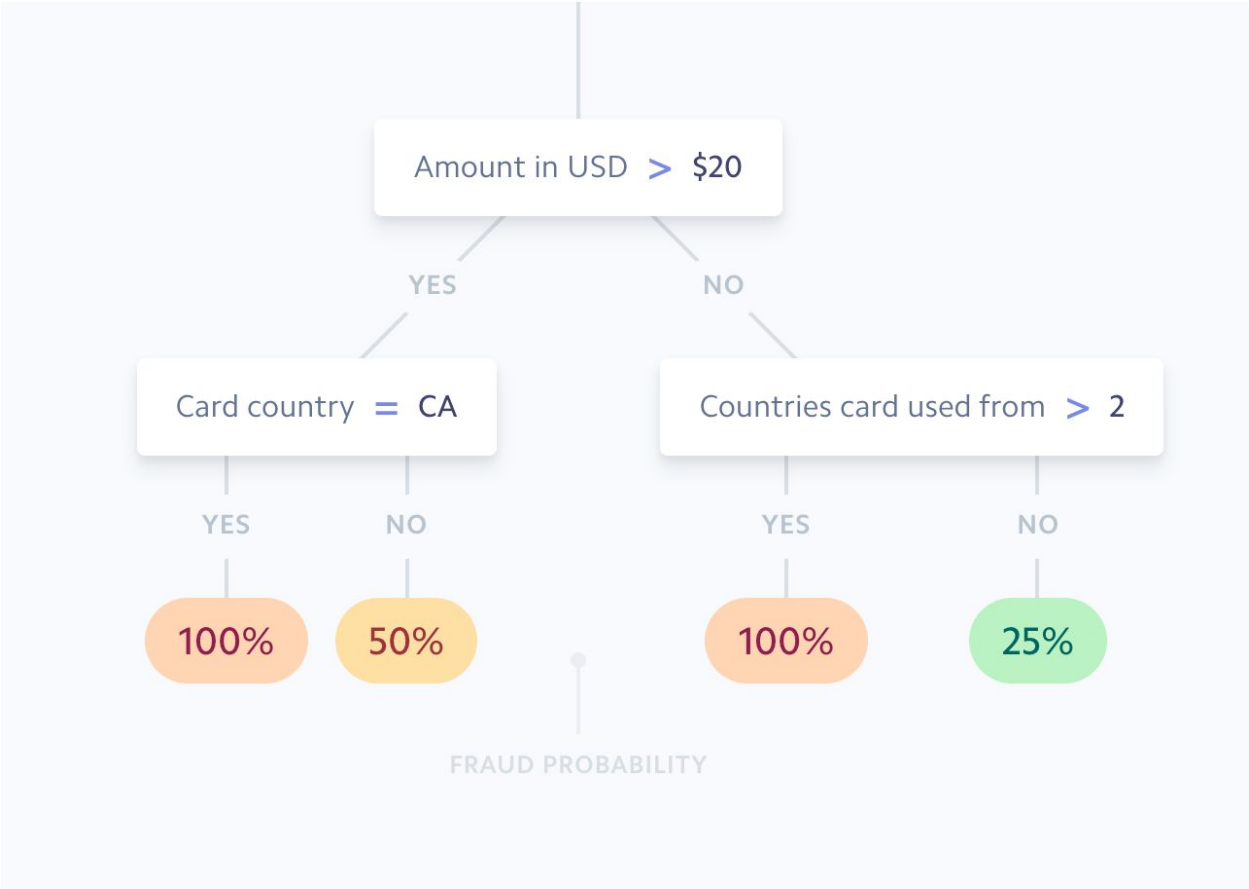
- Al tener datos de más de 100.000 empresas se tienen suficientes datos para utilizar la inteligencia artificial, lo cual no es viable para una empresa por sí misma por falta de datos.
- Al obtener los datos de fraude directamente del banco no necesitan analistas, así que evitan las subjetividades de marcar una transacción como fraudulenta o no (que quedaría reflejado en el modelo y podría impactar negativamente)

En este caso se utiliza el Aprendizaje automático para **predecir** un booleano: **true** si la transacción es fraudulenta y **false** en caso contrario. Se utilizan algunas variables de entrada como el país de la tarjeta de crédito, y el número de países desde el que se ha utilizado dicha tarjeta en las últimas 24 horas.

Los datos que se utilizan para **entrenar los modelos** se obtienen de transacciones (habitualmente históricos) de los que se conocen tanto los parámetros de entrada como la salida, como en el siguiente ejemplo:

Amount in USD	Card country	Countries card used from (24h)	Fraud?
$10.00	US	1	No
$10.00	CA	2	No
$10.00	CA	1	No
$10.00	US	1	Yes
$30.00	US	1	Yes
$99.00	CA	1	Yes
$15.00	CA	3	Yes
$70.00	US	1	No

Con miles de millones de transacciones entrenar un modelo puede resultar en el siguiente **árbol de decisiones:**



De esta manera pueden decidir cuándo bloquear o no una transacción fraudulenta.

Entrenamiento

Para poder **clasificar** una transacción como fraudulenta o no automáticamente, primero tenemos que haber visto muchos ejemplos de transacciones fraudulentas y sus **características.** Llamaremos características a las propiedades asociadas a una transacción relevantes para su clasificación, en el ejemplo anterior el país y el número de veces que ha sido utilizada la tarjeta. Necesitamos también un método para producir el modelo predictivo. En Stripe utilizan dos predicciones: la **regresión** y la **clasificación.** La primera da un valor numérico (el dinero perdido a causa de una transacción fraudulenta) y el segundo da un booleano (es fraude o no lo es). En el caso del boolean tendremos también un número asociado, que representa la probabilidad de ser fraude. Para la regresión se puede utilizar la regresión lineal o los árboles de regresión y para la clasificación árboles de decisión o bosques aleatorios, estos últimos utilizados actualmente en Stripe.

Evaluación de un modelo

Una vez entrenado un modelo, tenemos que comprobar cómo de eficiente es detectando el fraude. El modelo asigna una probabilidad o **puntuación P(fraude)** para decidir si es fraude o no una transacción. Llamaremos **F** a la función que si se cumple bloquea la transacción.

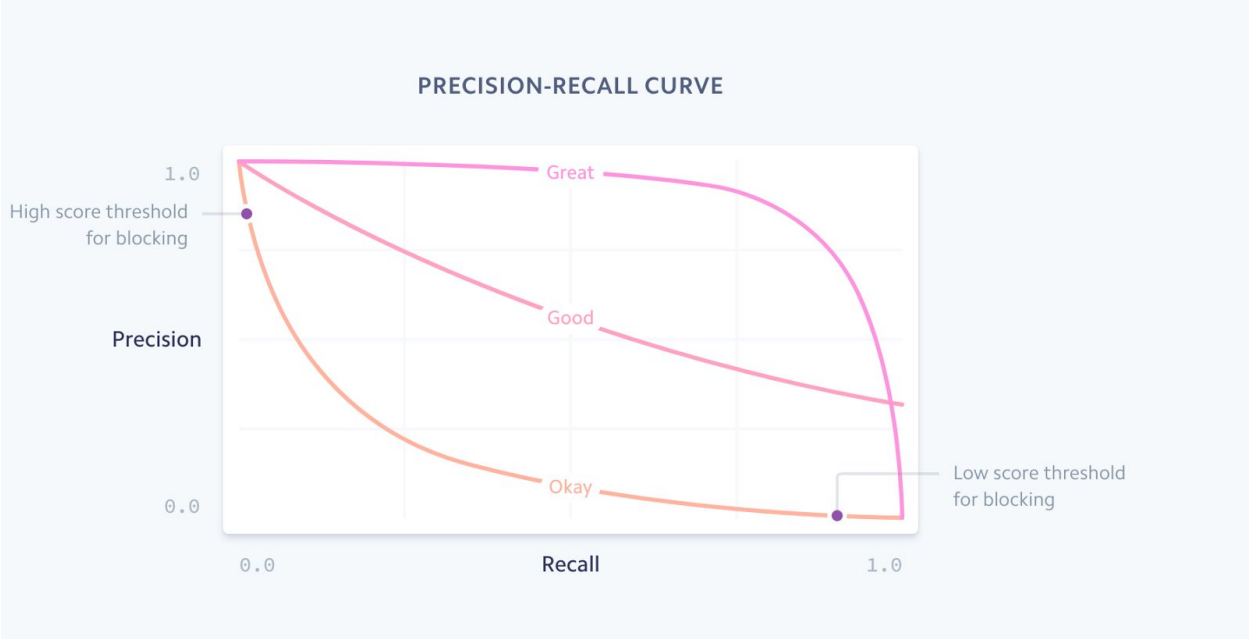
Terminología

Supongamos que bloqueamos pagos cuya **F = P(fraude) > 0.7,** para determinar la precisión de la función podemos utilizar varias piezas de información:

- **Precisión:** Es la fracción de transacciones que bloqueamos que son realmente fraudulentas, es decir que si bloqueamos 3 transacciones y 2 eran fraudulentas, la precisión será de **0.66**.
- **Sensibilidad (o recall):** La fracción de fraude que detectamos como tal. Si tenemos 5 transacciones fraudulentas y asignamos una **P(fraude) > 0.7** a 4 de ellas, tendremos que la sensibilidad es **0.8**.
- **Tasa de falsos positivos:** Es el número de pagos legítimos bloqueados erróneamente, es decir que si tenemos 10 pagos legítimos y bloqueamos 3 de ellos, la tasa de falsos positivos será de **0.3**.

Podemos deducir que unos valores óptimos de **precisión, sensibilidad y tasa de falsos positivos** serían **1.0, 1.0, 0.0** respectivamente, un modelo en el cual no tendríamos ningún falso positivo, todo el fraude es detectado y no bloqueamos ningún pago legítimo.

Podemos observar qué pasa variando los valores **X** de la función **F = P(fraude) > X** en la siguiente figura



Si asignamos una X demasiado grande tendremos mucha **precisión** ya que bloquearemos solo transacciones de las que estemos **muy convencidos**. Si asignamos una X demasiado baja tendremos mucha **sensibilidad,** ya que con mucha facilidad diríamos que una transacción es fraudulenta.

Buscamos un balance entre las dos, y esto es lo que se intenta optimizar en el aprendizaje automático.

Impacto en la empresa

El lanzamiento de Radar ha afectado a Stripe en los siguientes aspectos:

- Es mucho más eficiente en la detección del fraude que antes, ya que ahora ya no hay analistas que se dedicaban a programar reglas específicas.

- Ha supuesto importantes gastos (millones) a las empresas en que se utiliza, lo cual ha dado un importante empujón en su imagen y les ha supuesto una publicidad importante.

- Ha demostrado a la competencia que se puede usar el aprendizaje automático para combatir el fraude online. Otras empresas están investigando las posibilidades de utilizar este método.

Conclusión

El aprendizaje automático ha ayudado a Stripe a mejorar su negocio y han reafirmado el ámplio abanico de aplicaciones de la vida real que tiene esta técnica. Sin duda está en plena evolución y en los próximos años probablemente haya muchos avances, ya que las empresas más grandes del mundo están invirtiendo muchos recursos en ella.

Bibliografía

- https://en.wikipedia.org/wiki/Stripe_(company)
- http://fortune.com/2016/10/19/stripe-fraud-prevention/
- https://stripe.com/radar/guide
- http://neuralnetworksanddeeplearning.com/
- https://support.stripe.com/questions/radar-faq
- http://www.pymnts.com/news/security-and-risk/2016/stripe-radar-uses-machine-learning-to-block-fraud/