LICENCIATURA EN ESTADÍSTICA

DETECCIÓN DE FRAUDE

Un análisis con modelos lineales generalizados

Autores: Franco Santini - Nicolas Gamboa - Andrés Roncaglia

Docentes: Boggio Gabriela - Harvey Guillermina - Costa Vicotorio

2024

Tabla de contenidos

Introducción

El fraude con tarjetas de crédito es una de las principales amenazas que sufren los bancos. Con el auge de la tecnología las transacciones digitales facilitaron los traspasos de dinero y los medios de pago electrónicos son algo de cada día, pero junto con las ventajas también vinieron las consecuencias, y es que los métodos de fraude se han vuelto más sofisticados, generando pérdidas significativas a los bancos y afectando la confianza de los usuarios. Actividades como el uso no autorizado de tarjetas, la clonación de datos y transacciones fraudulentas requieren el desarrollo de tecnologías avanzadas para la detección temprana y la prevención.

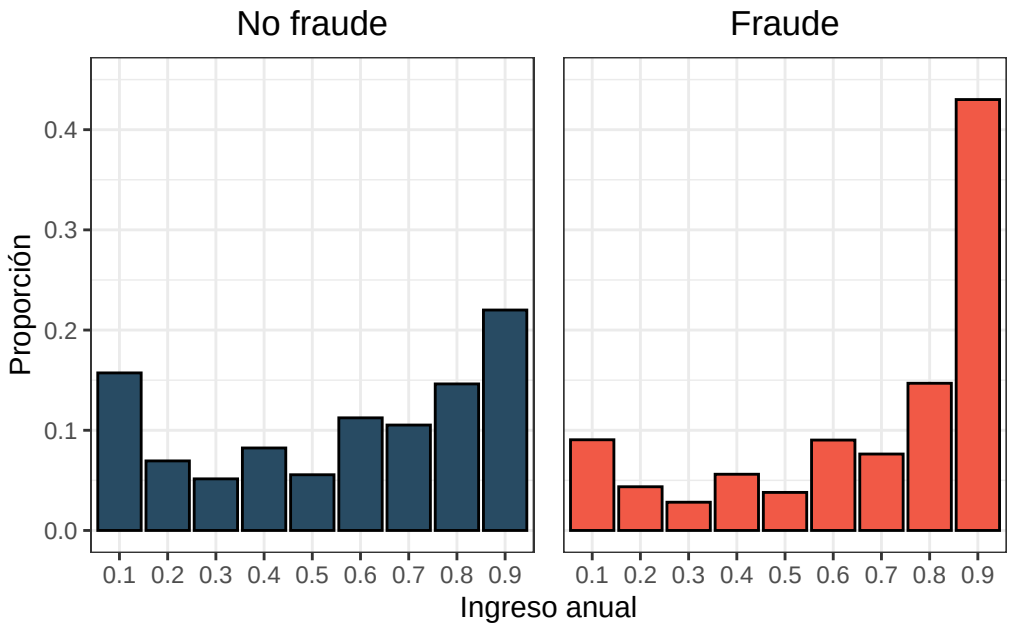
Variables:

- `fraud_bool`: Indicadora de si la transacción fue fraude o no
- `income`: Ingreso anual en cuantiles
- `name_email_similarity`: Similitud del nombre en el email y el nombre del solicitante
- `prev_address_months_count`: Es el número de meses que la persona estuvo viviendo en su locacion anterior
- `current_address_months_count`: Es el número de meses que la persona estuvo viviendo en su locacion actual
- `customer_age`: Edad del cliente en décadas
- `days_since_request`: Días desde la solicitud
- `intended_balcon_amount`: Valor de la transferecia inicial para aplicar al credito
- `payment_type`: Tipo del plan de pago
- `zip_count_4w`: Número de aplicaciónes con el mismo código postal en las últimas 4 semanas
- `velocity_6h`: Es la velocidad del total de solicitudes de transferencias de la tarjeta en las últimas 6 horas
- `velocity_24h`: Es la velocidad del total de solicitudes de transferencias de la tarjeta en las últimas 24 horas
- `velocity_4w`: Es la velocidad del total de solicitudes de transferencias de la tarjeta en las últimas 4 semanas
- `bank_branch_count_8w`: Número total de solicitudes en la seleccionada rama del banco en las últimas 8 semanas
- `date_of_birth_distinct_emails_4w`: Número de emails de aplicantes con la misma fecha de nacimiento en las últimas 4 semanas
- `employment_status`: Estado de empleo del solicitante
- `credit_risk_score`: Score de riesgo de la aplicación
- `email_is_free`: Tipo del dominio del email del aplicante (email pago o gratis)

- `housing_status`: Estado residencial del aplicante
- `phone_home_valid`: Validez del telefono fijo provisto
- `phone_mobile_valid`: Validez del telefono movil provisto
- `bank_months_count`: Antiguedad de la cuenta anterior en meses
- `has_other_cards`: Indicador de si la persona tiene otra tarjeta en el mismo banco
- `proposed_credit_limit`: Crédito limite propuesto por el aplicante
- `foreign_request`: Indicadora de si la solicitud fue hecha en el mismo pais que el banco
- `source`: Fuente online de la aplicación (Internet / app movil)
- `session_length_in_minutes`: Tiempo de la sesion en la pagina del banco en minutos
- `device_os`: Sistema operativo del dispositivo que hizo la solicitud
- `keep_alive_session`: Indicadora de si el solicitante decidió mantener la sesión iniciada al ingresar
- `device_distinct_emails_8w`: Número de emails distintos en la página del banco desde el mismo dispositivo usado en las últimas 8 semanas
- `device_fraud_count`: Número de solicitudes fraudulentas desde el dispositivo utilizado
- `month`: Mes en el que fue realizada la solicitud

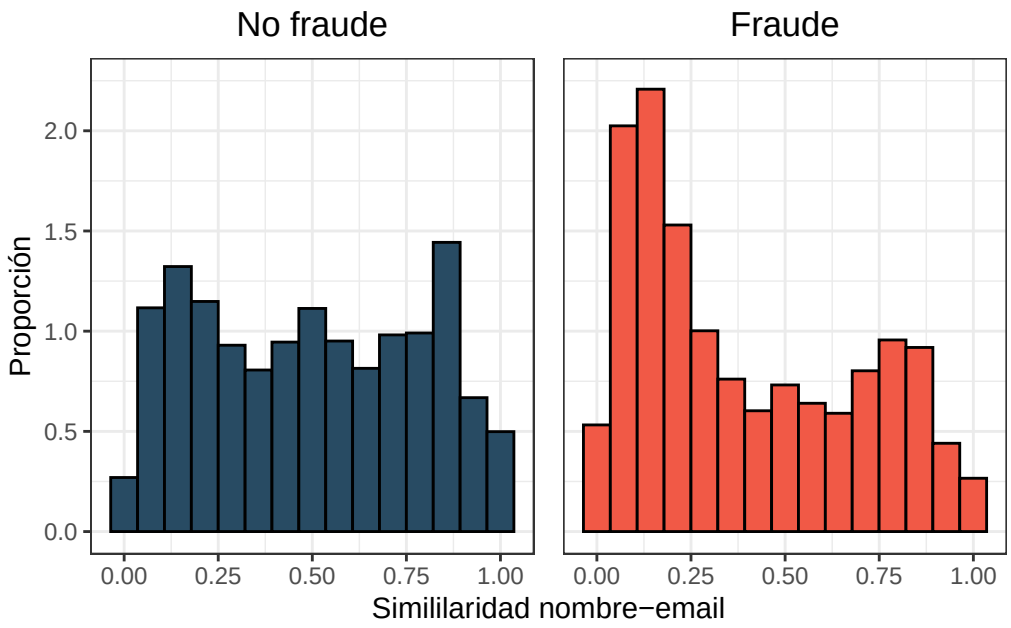
Análisis descriptivo

Figura 1: Distribución del ingreso anual según si la transacción es fraudulenta o no



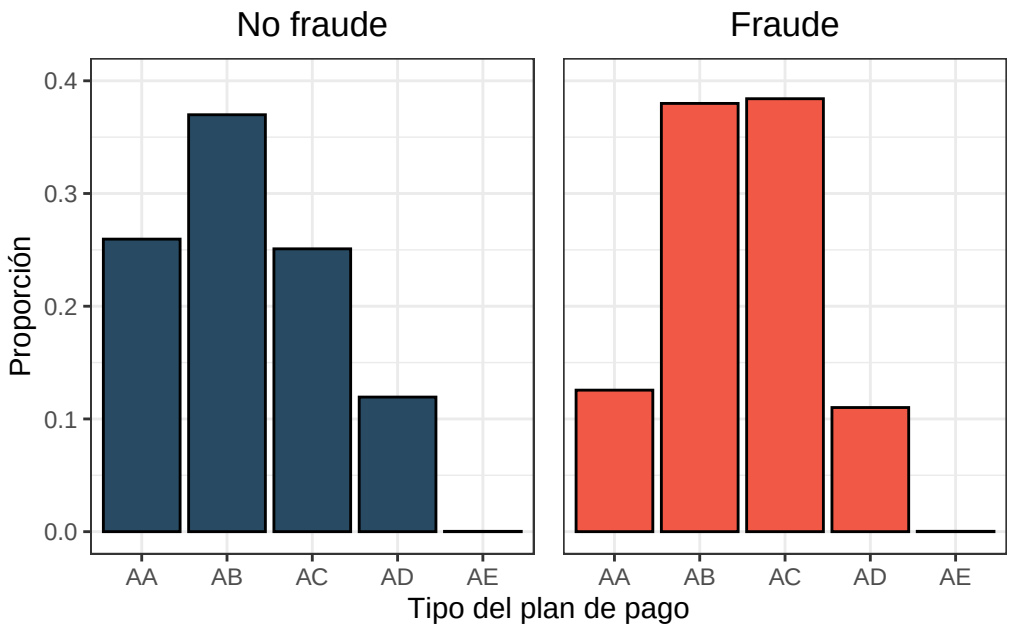
Se puede observar que las personas que cometieron fraude tienden a tener un ingreso anual registrado mayor. La distribución tiene una mayor asimetría a la izquierda.

Figura 2: Distribución del índice de similitud entre en nombre del solicitante y el nombre en el email según si la transacción es fraudulenta o no



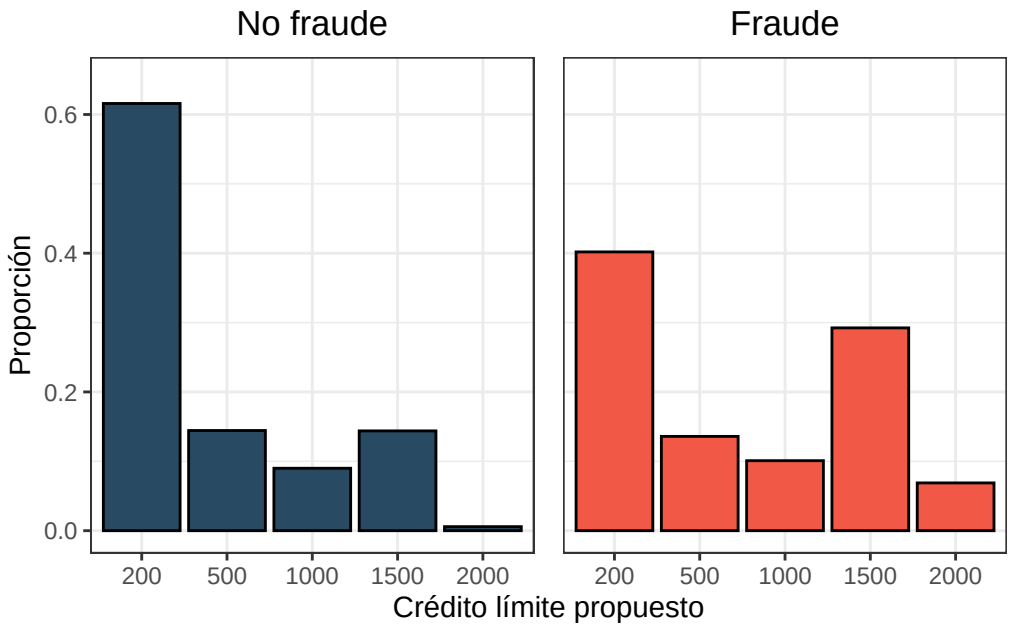
Las transacciones realizadas con emails no muy similares al nombre real de la persona parecen ser más propensas a ser fraudulentas.

Figura 3: Proporción del tipo de pago según si la transacción es fraudulenta o no



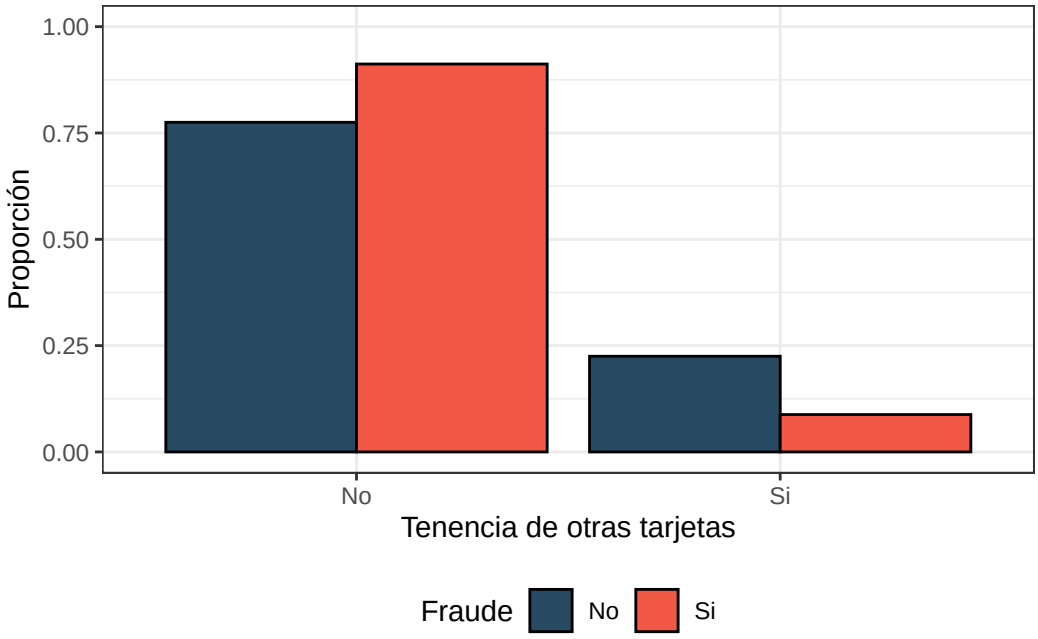
En general, las personas que cometen fraude parecen preferir los métodos de pago “AB” y “AC” por encima del resto, al contrario de las personas que operan de manera legítima que prefieren de igual manera los tipos de pago “AA”, “AB” y “AC”. Se puede notar también que la forma de pago “AE” no es muy popular.

Figura 4: Distribución del límite crediticio propuesto por el solicitante según si la transacción es fraudulenta o no



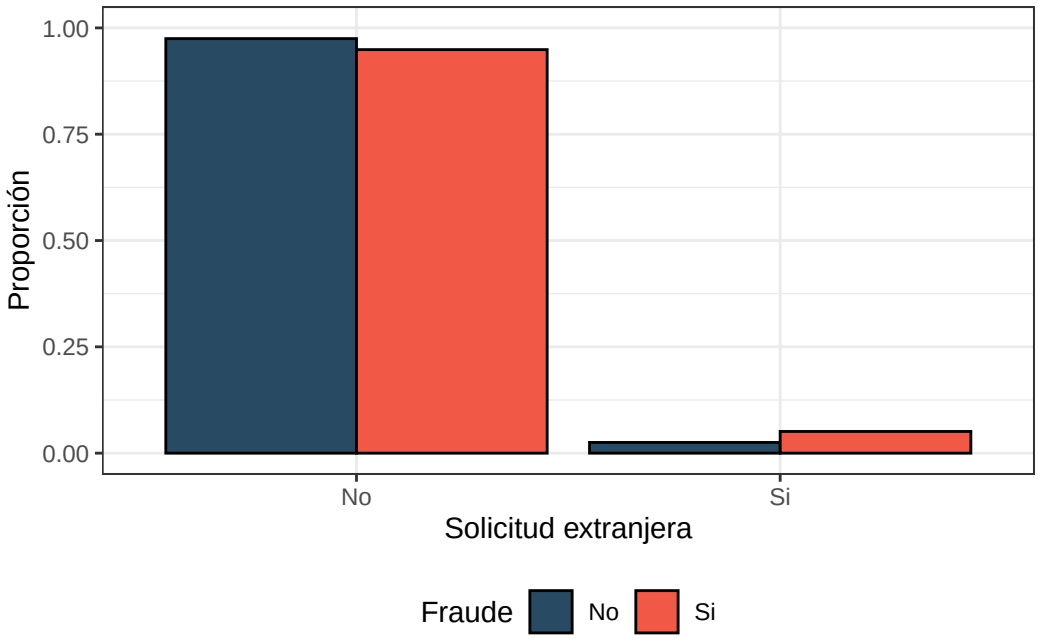
Se puede destacar en este gráfico que las personas que cometen fraude son ligeramente más propensas a pedir créditos más altos.

Figura 5: Proporción de la tenencia de otra tarjeta en el mismo banco según si la transacción es fraudulenta o no



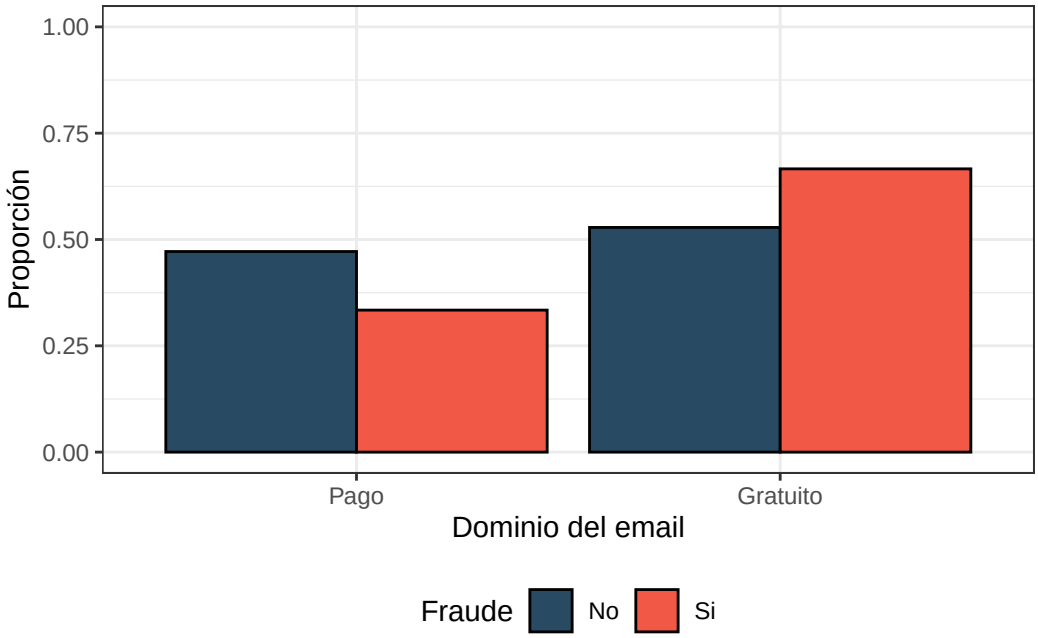
En cuanto a la tenencia de otra tarjeta en el mismo banco, suele ser común no poseer otra, sin embargo las personas que operan de forma legal se inclinan a tener más de una tarjeta un poco más que aquellos que cometen fraude.

Figura 6: Proporción de la locación de la solicitud según si la transacción es fraudulenta o no



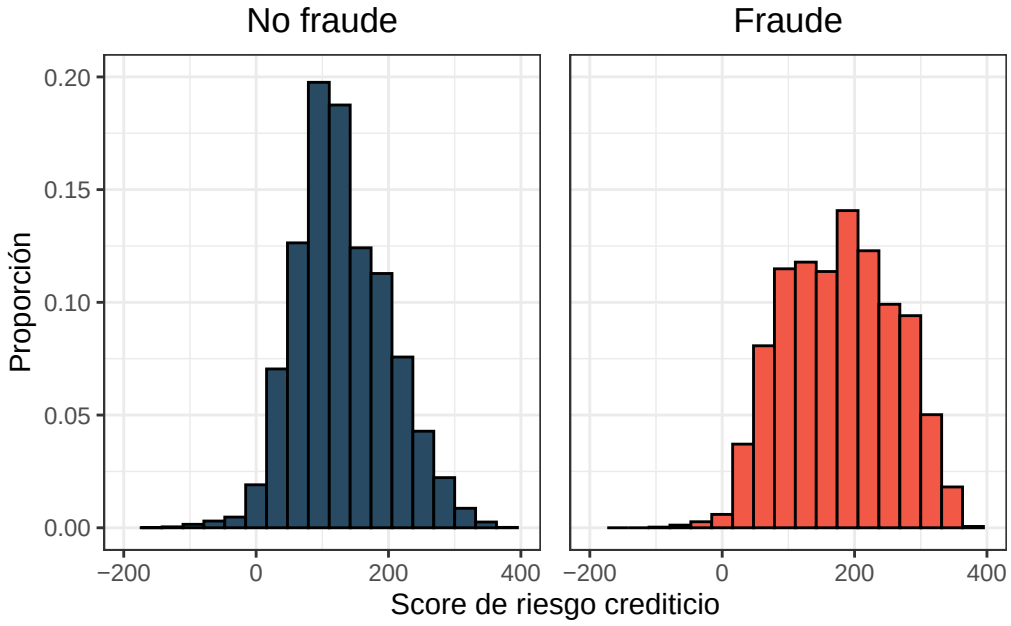
Se puede observar que las personas que cometen fraude, parecen hacer más solicitudes del exterior que las personas que no cometen fraude, aunque la diferencia parece ser sutil.

Figura 7: Proporción del tipo de dominio del email según si la transacción es fraudulenta o no



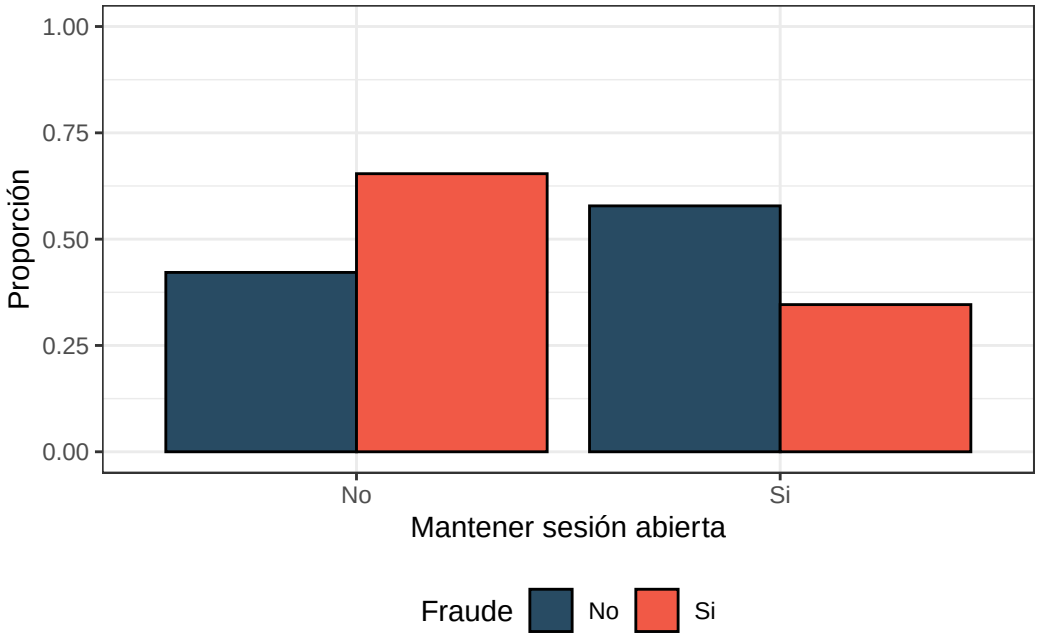
También se puede destacar que las operaciones fraudulentas parecen ser más comunes cuando el dominio del email del solicitante es gratuito que cuando es pago.

Figura 8: Distribución del score de riesgo interno según si la transacción es fraudulenta o no



Lla distribución del score de riesgo para las personas que cometen fraude es simétrica y centrada alrededor de 200, mientras que la distribución del score de riesgo para las personas que no cometen fraude parece ser más asimétrica y tener una media menor.

Figura 9: Proporción de opciones de inicio de sesión según si la transacción es fraudulenta o no



Por lo general, cuando las transacciones son fraudulentas la persona decide no mantener la sesión abierta en la cuenta del banco en mayor proporción que cuando las transacciones son legítimas.