

TRANSACCIONES FRAUDULENTAS

Victoria Vallejo
14/09/2022

CONTENIDOS

01

ACTUALIDAD

Transacciones fraudulentas en el presente.

02

EDA

Exploramos la base de datos.

03

MODELOS

- Árbol de Decisión
- Reg Logística

04

CONCLUSIÓN

¿Qué modelo elegimos?

05

RECOMENDACIONES

En base a los resultados

01

ACTUALIDAD



En el primer semestre del año, la Condusef detectó 3.5 millones de fraudes cibernéticos y los negocios en internet, sin importar el tamaño, son uno de los sectores más afectados.

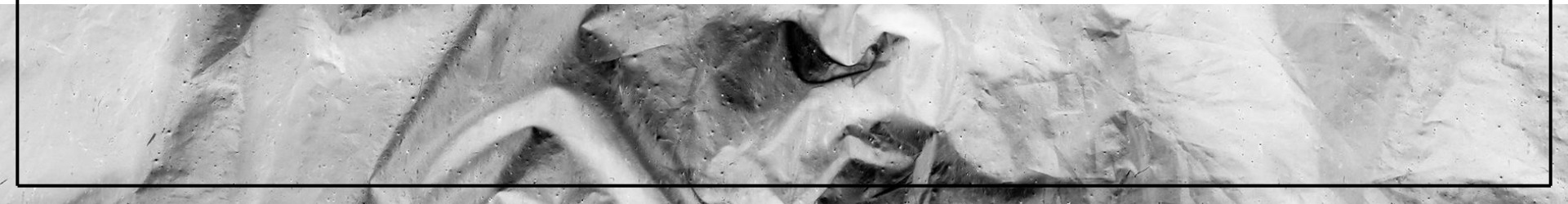
El fraude bancario aumenta un 159% y las transacciones alcanzan los volúmenes anteriores a la pandemia

Entre la espada y la pared: una nueva investigación revela que la banca de EE. UU. está atrapada por el aumento de los delitos financieros y la complejidad tecnológica percibida

El fraude con tarjetas de crédito resultó en la pérdida de \$ 3 mil millones para las instituciones financieras de América del Norte en 2017. El auge de los sistemas de pagos digitales como Apple Pay, Android Pay y Venmo ha significado que se espera que aumenten las pérdidas debido a actividades fraudulentas. Deep Learning presenta una solución prometedora al

159%

AUMENTO



02

EDA

DATASET

Conjunto de datos sintéticos generados usando el simulador PaySim. PaySim utiliza datos agregados del conjunto de datos privados para generar un conjunto de datos sintéticos que se asemeja a la operación normal de las transacciones.

6362620

Registros

11

Variables

0

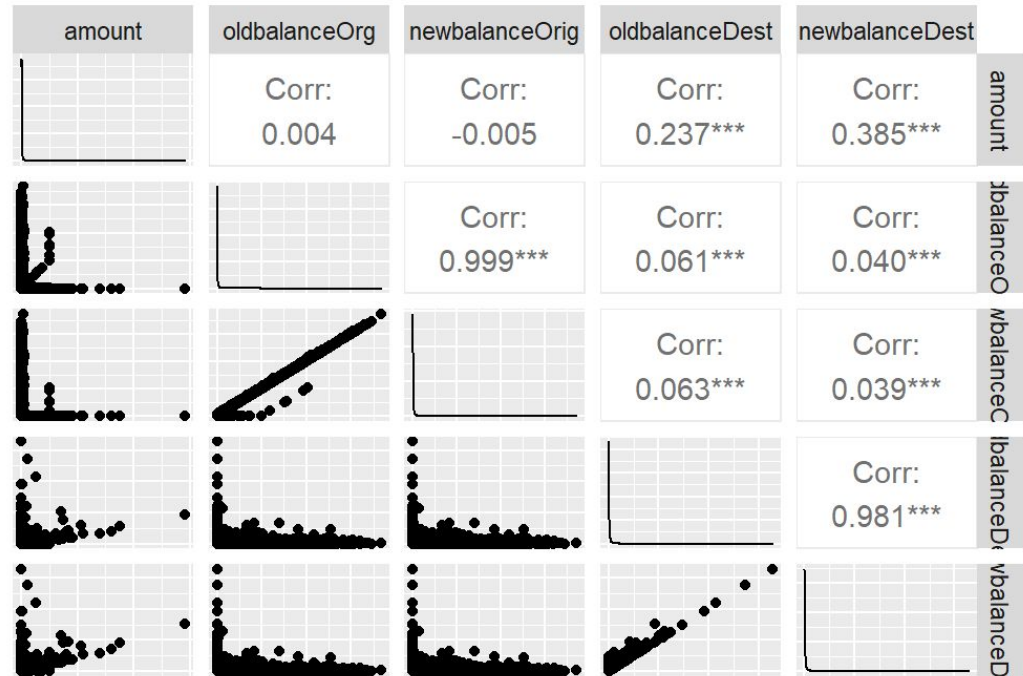
NULL values

VARIABLES

```
'data.frame': 6362620 obs. of 11 variables:
 $ step      : Factor w/ 743 levels "1","2","3","4",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ type      : Factor w/ 5 levels "CASH_IN","CASH_OUT",...: 4 4 5 2 4 4 4 4 4 3 ...
 $ amount    : num  9840 1864 181 181 11668 ...
 $ nameOrig  : chr   "C1231006815" "C1666544295" "C1305486145" "C840083671" ...
 $ oldbalanceOrig : num  170136 21249 181 181 41554 ...
 $ newbalanceOrig: num  160296 19385 0 0 29886 ...
 $ nameDest   : chr   "M1979787155" "M2044282225" "C553264065" "C38997010" ...
 $ oldbalanceDest: num   0 0 0 21182 0 ...
 $ newbalanceDest: num   0 0 0 0 0 ...
 $ isFraud    : Factor w/ 2 levels "0","1": 1 1 2 2 1 1 1 1 1 1 ...
 $ isFlaggedFraud: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
```

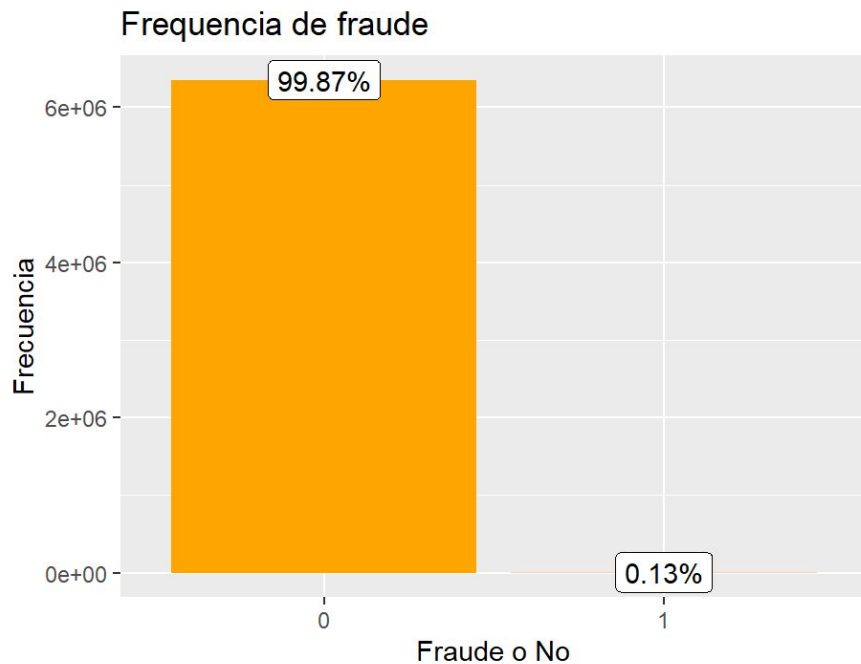

CORRELACIÓN

Numeric Features

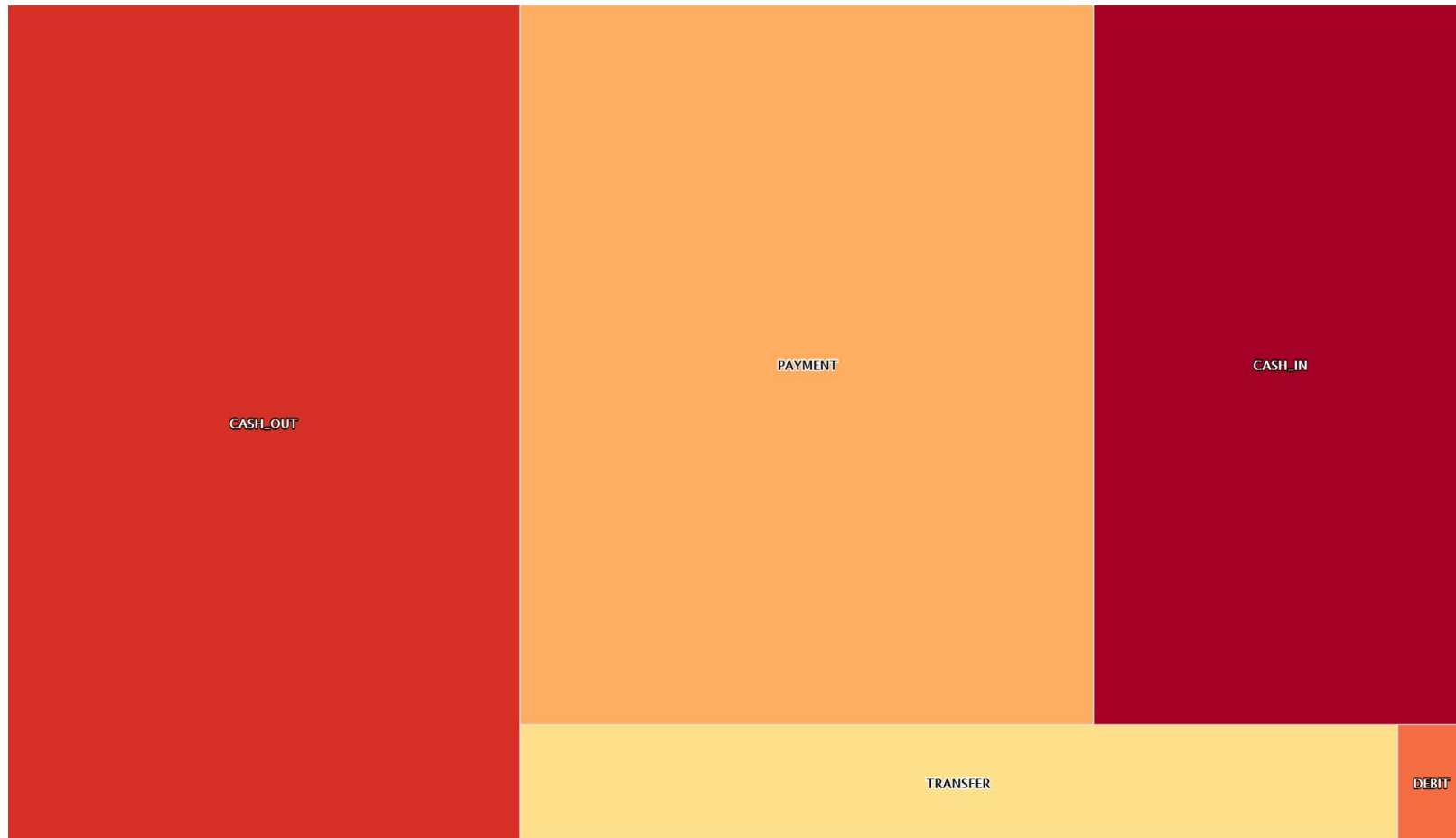


FRAUDES EN LA BASE

isFraud	cant	freq
<fct>	<int>	<dbl>
0	6354407	0.999
1	8213	0.00129



Cantidad de transacciones por tipo de medio

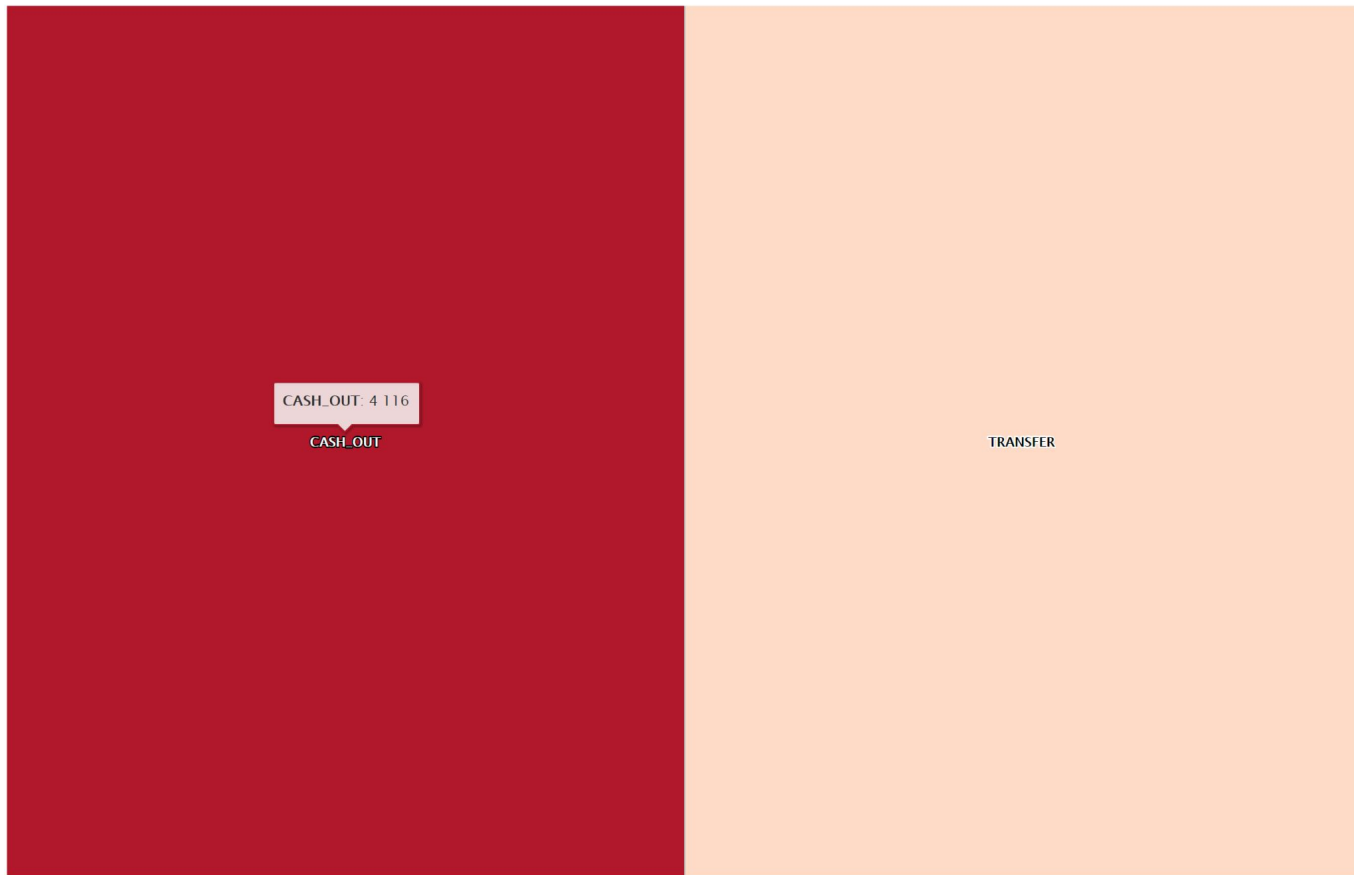


8213

REGISTROS DE FRAUDES

DE FRAUDES

Transacciones fraude por tipo de medio



Transacciones fraude por tipo de medio



No hay fraudes del tipo "CASH_IN", "DEBIT", "PAYMENT" !

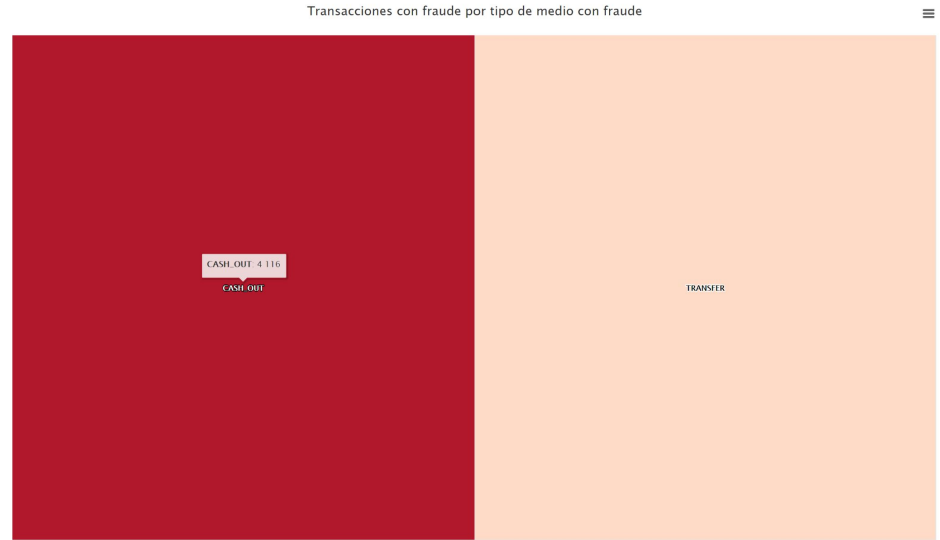
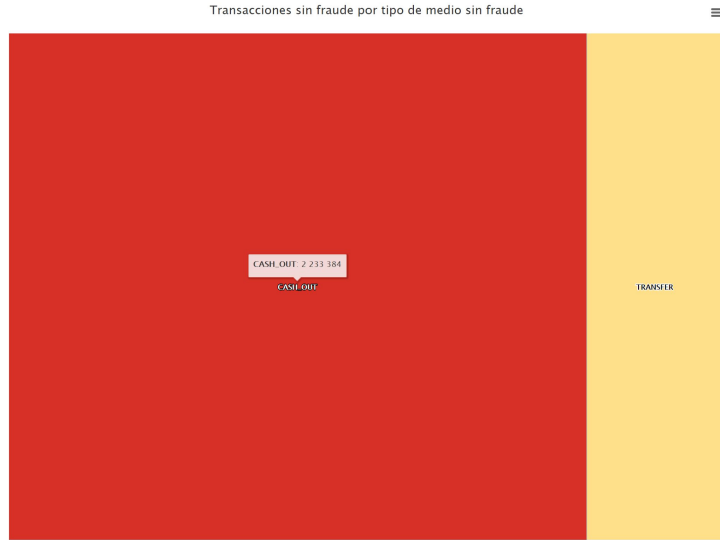
La variable “type” seguramente sea significativa para la clasificación de la variable “isFraud”.

Si se eliminan sólo para el siguiente gráfico los registros con “type”= "CASH_IN";"DEBIT";"PAYMENT" quedarían

2770409

11

SIN FRAUDE / CON FRAUDE

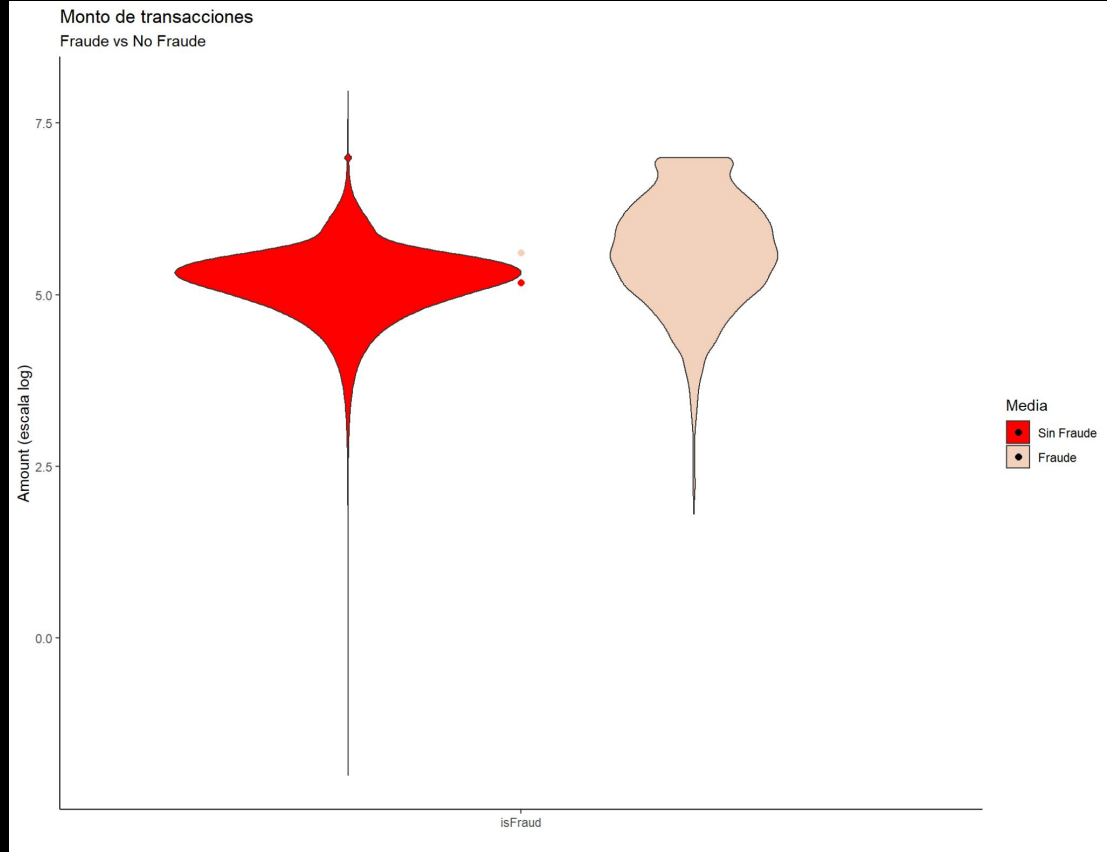


Los registros sin fraude presentan en su mayoría “cash_out” (retiro efectivo), mientras que en los registros con fraude la frecuencia de “cash_out” y “transfer” es casi idéntica lo cuál podría ser útil a la hora de predecir fraude.

AMOUNT (LOG10)

Fraude: El monto muestra una distribución más homogénea que sin fraude.
Sin fraude: El monto se encuentra frecuentemente en 100000

Amount será una variable significativa a la hora de predecir isFraud.



CHI Q TEST

Que sea fraude o no (isFraud) es independiente de tener un monto (amount) mayor a 100000?

H0: fraude \leq 100000

Ha: fraude $>$ 100000

```
a = table(df$isFraud,df$amount>100000) # se discretiza la variable numerica  
chisq.test(a)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: a  
X-squared = 3988.3, df = 1, p-value < 2.2e-16
```

P-value $< 0.05 \rightarrow$ el modelo es significativo a un 95% de confianza por lo tanto rechazamos H0.
X-squared es lejano al 0, no son independientes.

03

MODELOS

SAMPLE

Al presentar una gran cantidad de registros R no corre. Fue necesario reducir los registros a un 10% para poder realizar los modelos.

Los modelos fueron realizados con 636261 registros de la base.

ÁRBOL DE DECISIÓN

Partición en Train y Test

```
set.seed(8) ; particion= createDataPartition(y=df_numc$isFraud, p=0.8,  
list=FALSE)
```

```
Train ← df_numc[particion,]
```

```
Test ← df_numc[-particion,]
```

ÁRBOL DE DECISIÓN

P = 80%

MATRIZ DE CONFUSIÓN

Prediction / Reference	No Fraude	Fraude
No Fraude	635413	337
Fraude	24	487

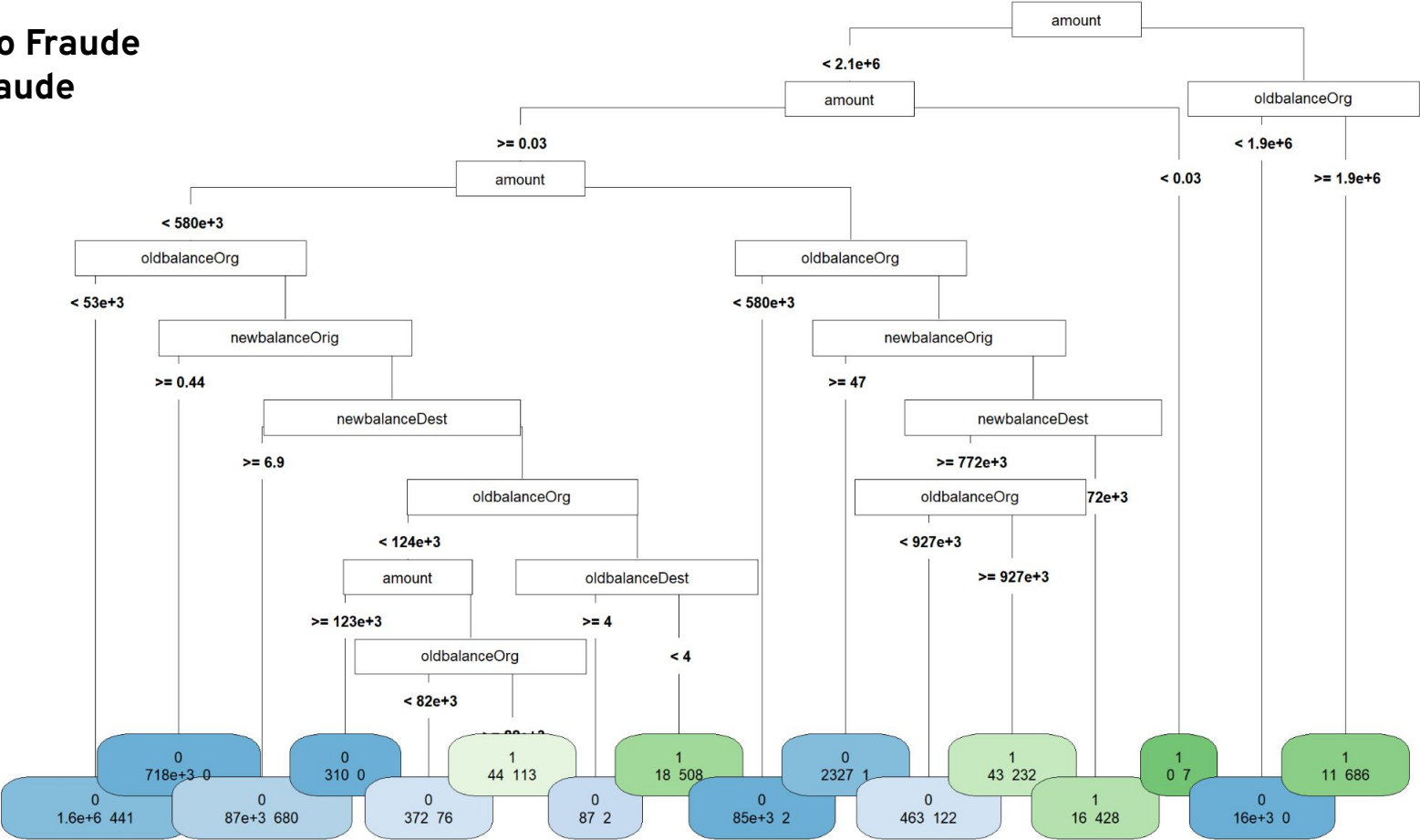
ÁRBOL DE DECISIÓN

P = 80%

TEST

ACCURACY : 0.9994 = 99.94%

0 No Fraude
1 Fraude



REGRESIÓN LOGÍSTICA

$P = 80\%$

REGRESIÓN LOGÍSTICA

MATRIZ DE CONFUSIÓN

Prediction / Reference	No Fraude	Fraude
No Fraude	634058	404
Fraude	847	2070

REGRESIÓN LOGÍSTICA

TEST

ACCURACY : 0.9980373 = 98.80%

04 CONCLUSIÓN

MODELO

Las variables “type” y “amount” son significativas para predecir el fraude.

El modelo más efectivo para utilizar al predecir fraudes es el Árbol de decisión con una accuracy del **99.94%**

05

RECOMENDACIONES

EN BASE A LO ANALIZADO SE RECOMIENDA ...

1. Implementar mayores medidas de seguridad para transacciones del tipo cash_out (retiro efectivo) y transfer (transferencia), donde se encuentran las transacciones fraudulentas por ejemplo mediante autenticación obligatoria de dos pasos y biometría.



EN BASE A LO ANALIZADO SE RECOMIENDA ...

2. Reforzar la seguridad de la plataforma bancaria.



EN BASE A LO ANALIZADO SE RECOMIENDA ...

3. Establecer y comunicar de un único medio de contacto oficial con los clientes por parte del banco para evitar ataques phishing (el atacante envía un correo electrónico o SMS con un enlace para que la persona entre a la cuenta para solucionar algún problema o llevar a cabo una acción necesaria)



EN BASE A LO ANALIZADO SE RECOMIENDA ...

4. Educar al cliente sobre los pasos a seguir una vez identificado un fraude bancario.



EN BASE A LO ANALIZADO SE RECOMIENDA ...

5. Campaña “Tome medidas para protegerse” donde se alentará al cliente a:

- Habilitar la biometría (inicio de sesión con huellas dactilares o reconocimiento facial)
- Mantener su información de contacto actualizada
- Fortalecer sus contraseñas
- Activar las alertas automáticas en la aplicación del banco.



GRACIAS!

