

Fraudulent Transactions



Seminario Profesional I

Brandon Garcia - 19005926

Gabriel Monzon - 19000437

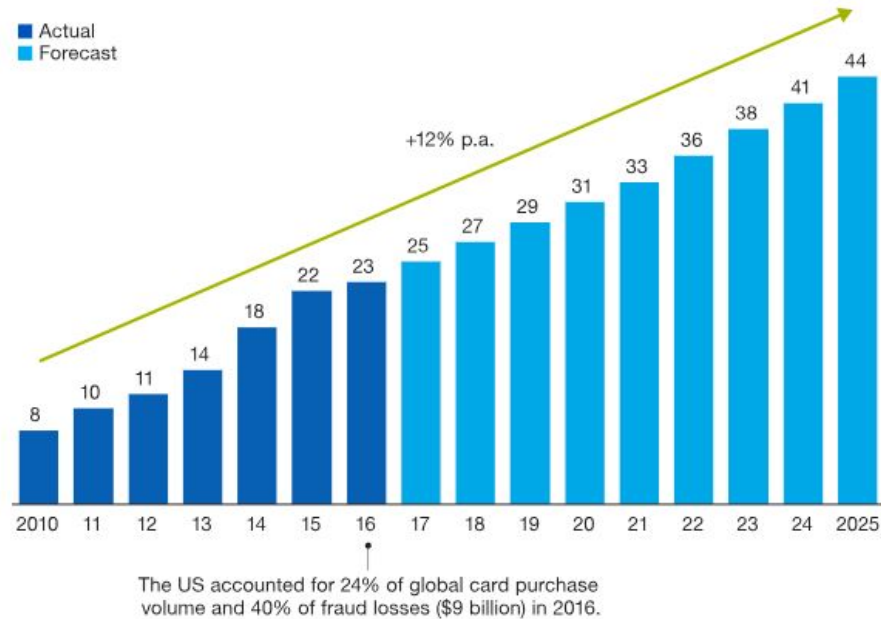
Pedro Cuá - 19001516

Problema

Transacciones Fraudulentas

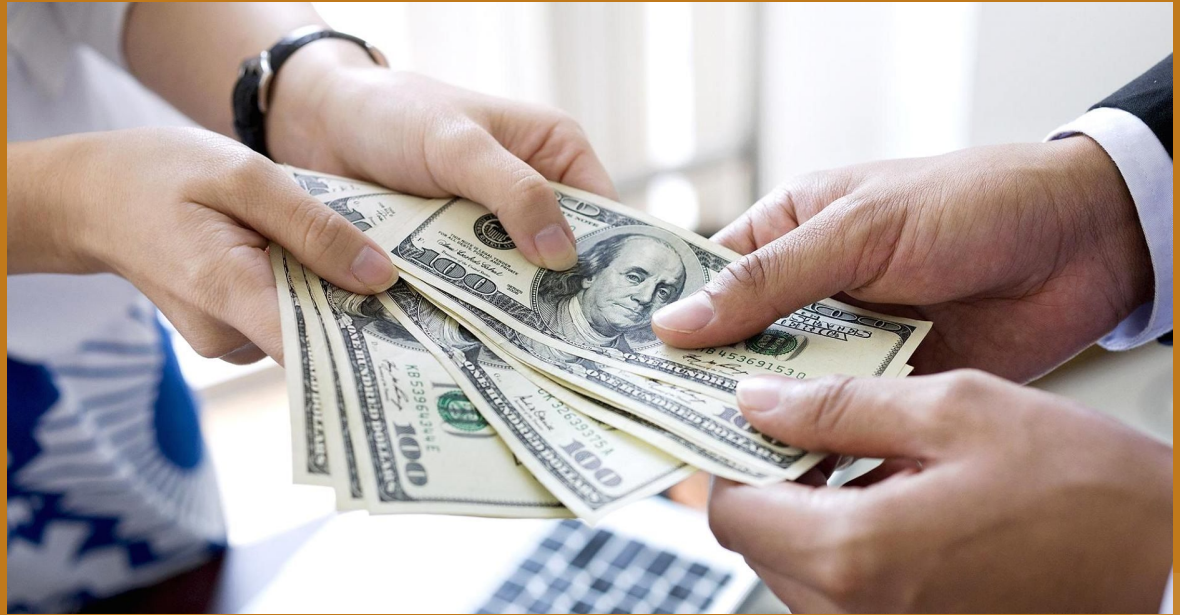


Global card fraud losses totaled almost \$23 billion in 2016.¹



¹General-purpose and private-label global and domestic credit, debit, and prepaid cards; excludes operational costs incurred by issuers, merchants, acquirers, call centers, and chargeback management.

Cash In



Cash Out



Payment



Transfer



Beneficio e Importancia de la Solución



**COMPROMISED
CUSTOMER DEVICE**



**LEGITIMATE
CUSTOMER**



**IDENTITY
THEFT**



RISK & PROTECTION SERVICES



ALGORITHMIC
AWARENESS



REAL-TIME
ANALYTICS



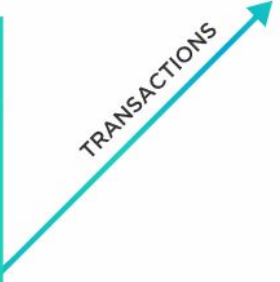
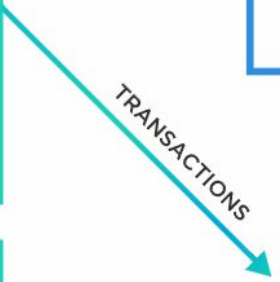
CUSTOMER DATA
MANAGEMENT



FRONT OFFICE



BACK OFFICE





Conjunto de datos

Variables

*step: Representa una hora.

type: Tipo de transacción fraudulenta.

amount: Dinero de la transacción fraudulenta.

*nameOrig: Cliente que realizó la transacción.

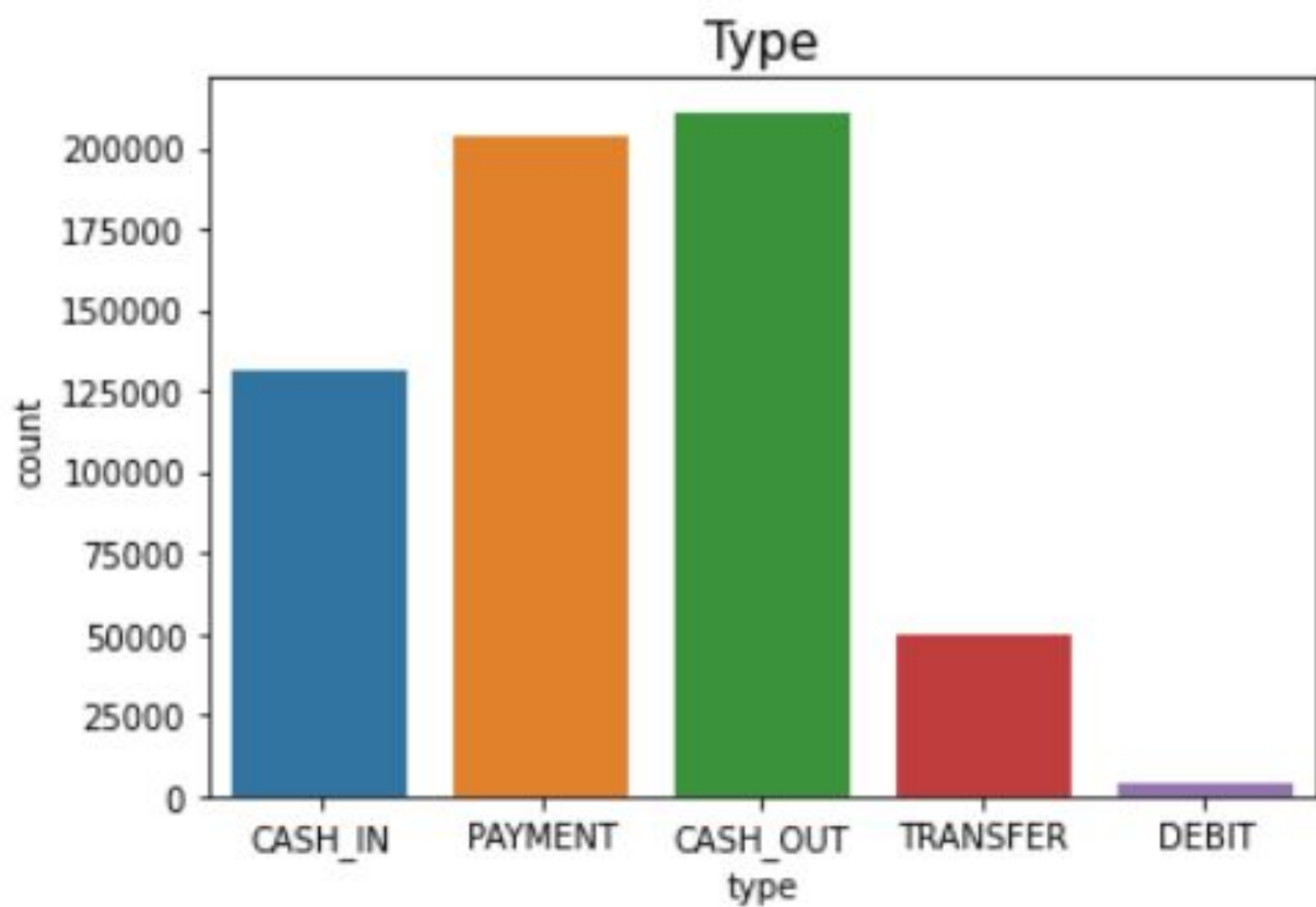
oldbalanceOrg: Dinero de la cuenta origen antes de la transferencia.

newbalanceOrig: Dinero de la cuenta origen después de la transferencia.

*nameDest: Cliente que recibió la transacción.

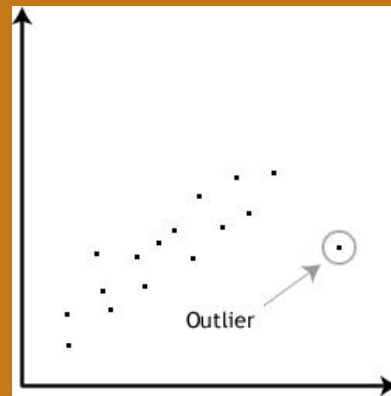
oldbalanceDest: Dinero de la cuenta destino antes de la transferencia.

newbalanceDest: Dinero de la cuenta destino después de la transferencia.



¿Cómo detectar transacciones Fraudulentas?

Deteccion de anomalias o patrones inusuales que son distintos a los comportamientos esperados.





Arquitectura



input_1	input:	[(None, 6)]	[(None, 6)]
InputLayer	output:		



dense	input:	(None, 6)	(None, 128)
Dense	output:		



dense_1	input:	(None, 128)	(None, 128)
Dense	output:		



dense_2	input:	(None, 128)	(None, 128)
Dense	output:		



dense_3	input:	(None, 128)	(None, 128)
Dense	output:		



dense_4	input:	(None, 128)	(None, 2)
Dense	output:		

Autoencoders

6

5

4

3

4

5

6

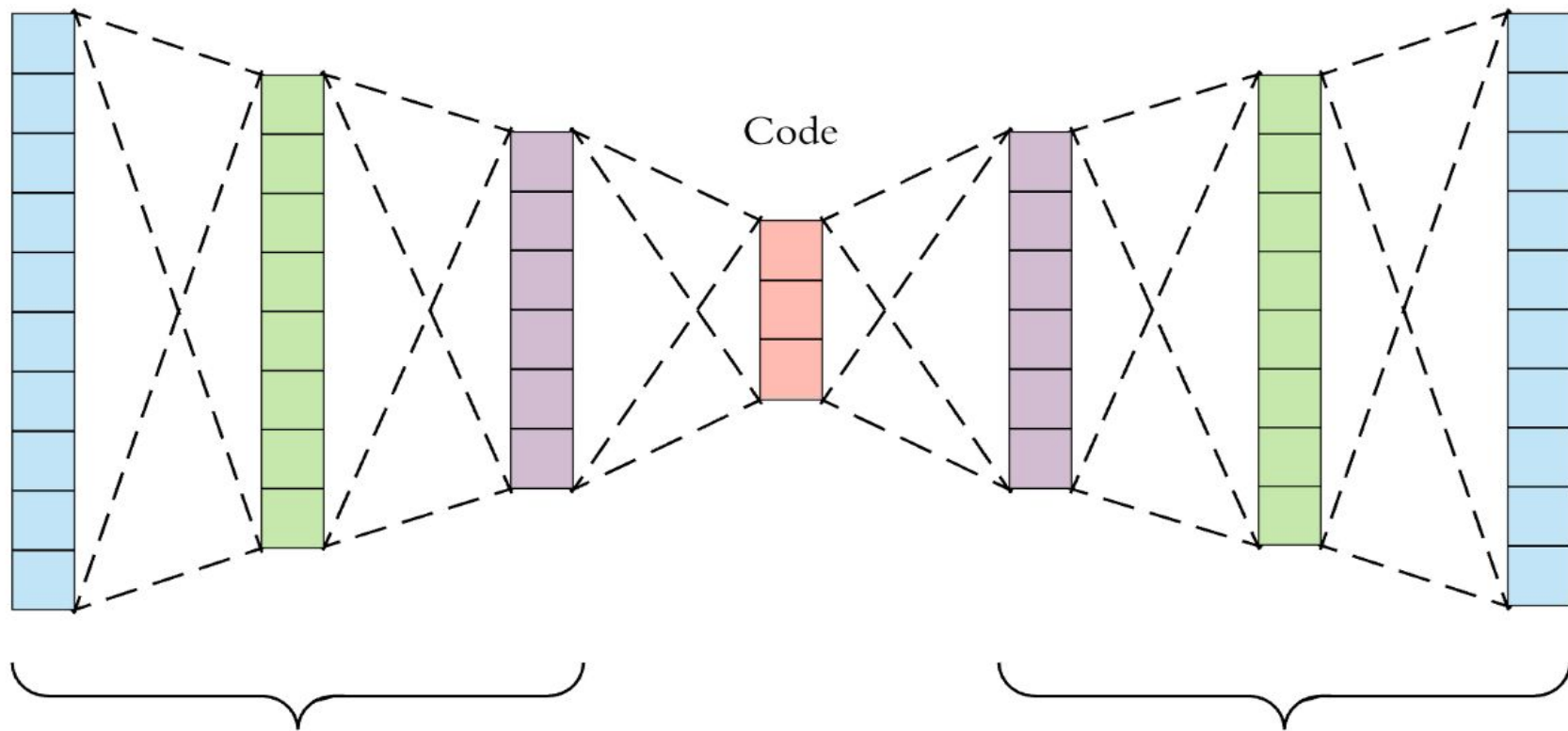
Input

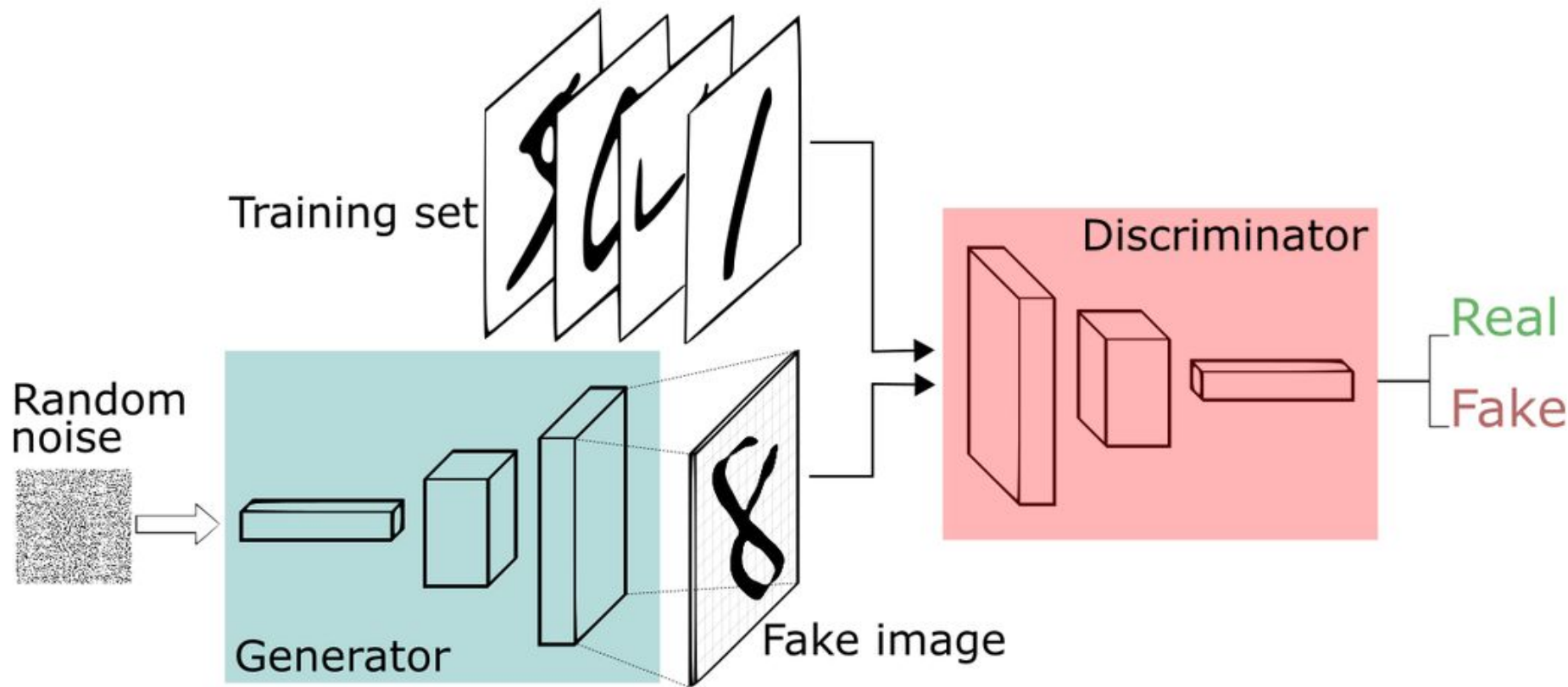
Output

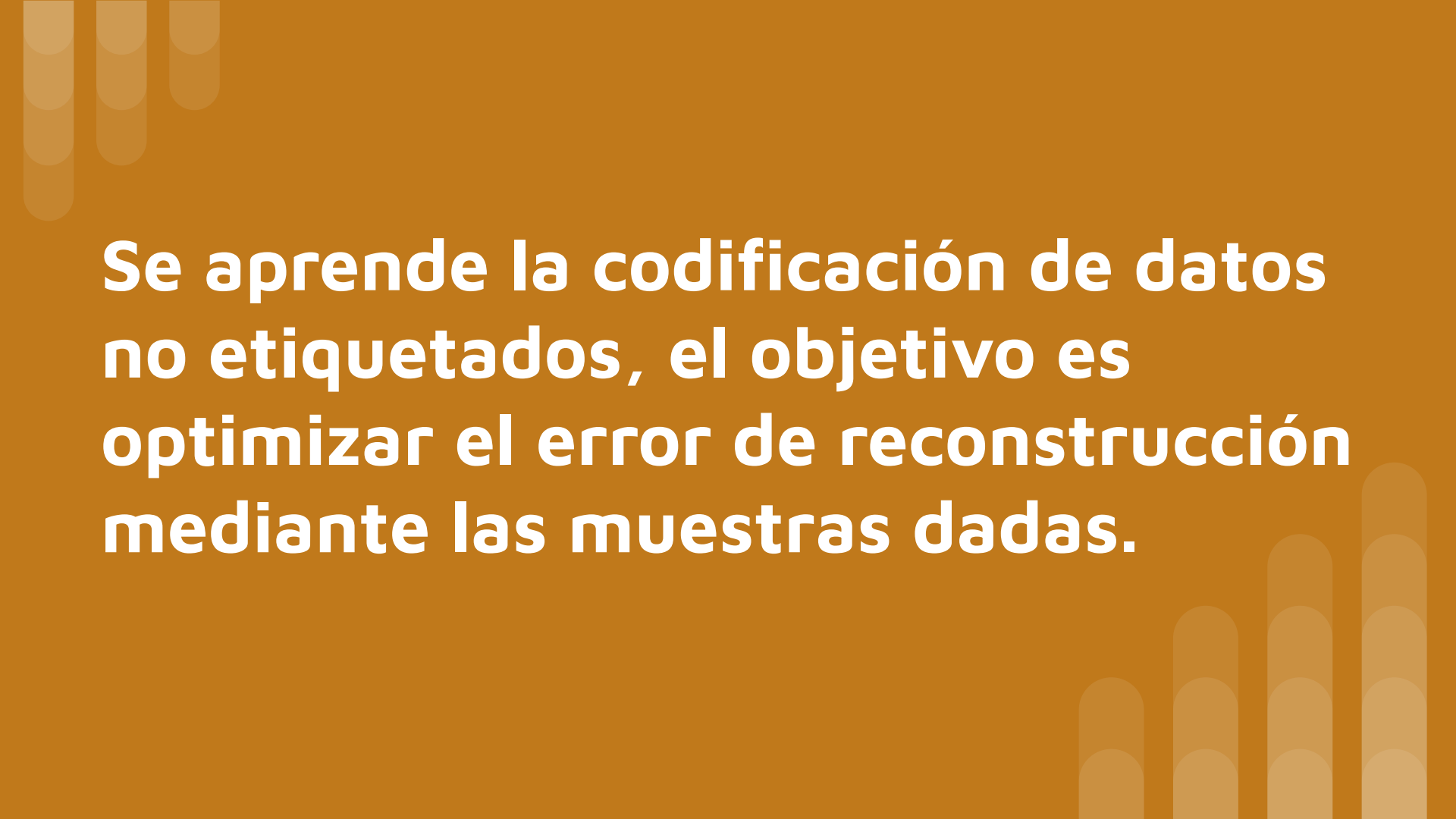
Code

Encoder

Decoder







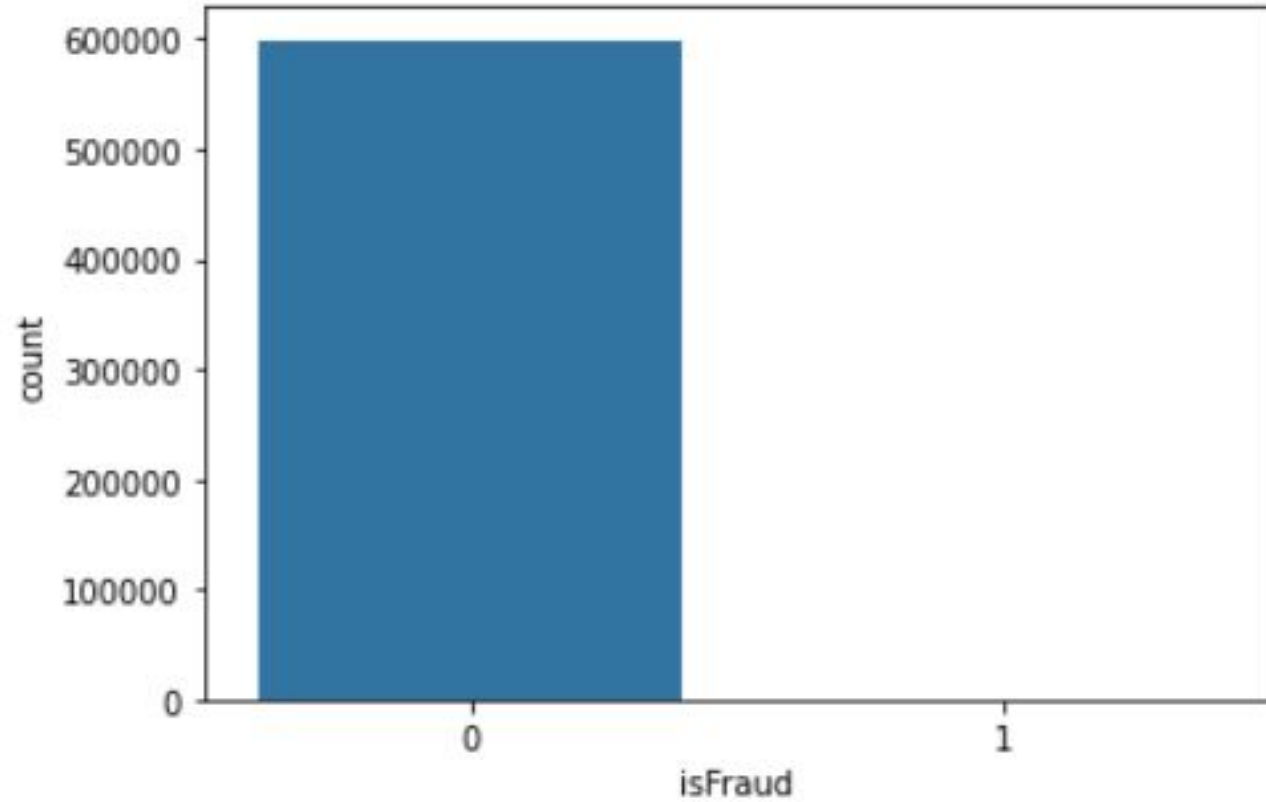
Se aprende la codificación de datos no etiquetados, el objetivo es optimizar el error de reconstrucción mediante las muestras dadas.



Experimentación



Fraud



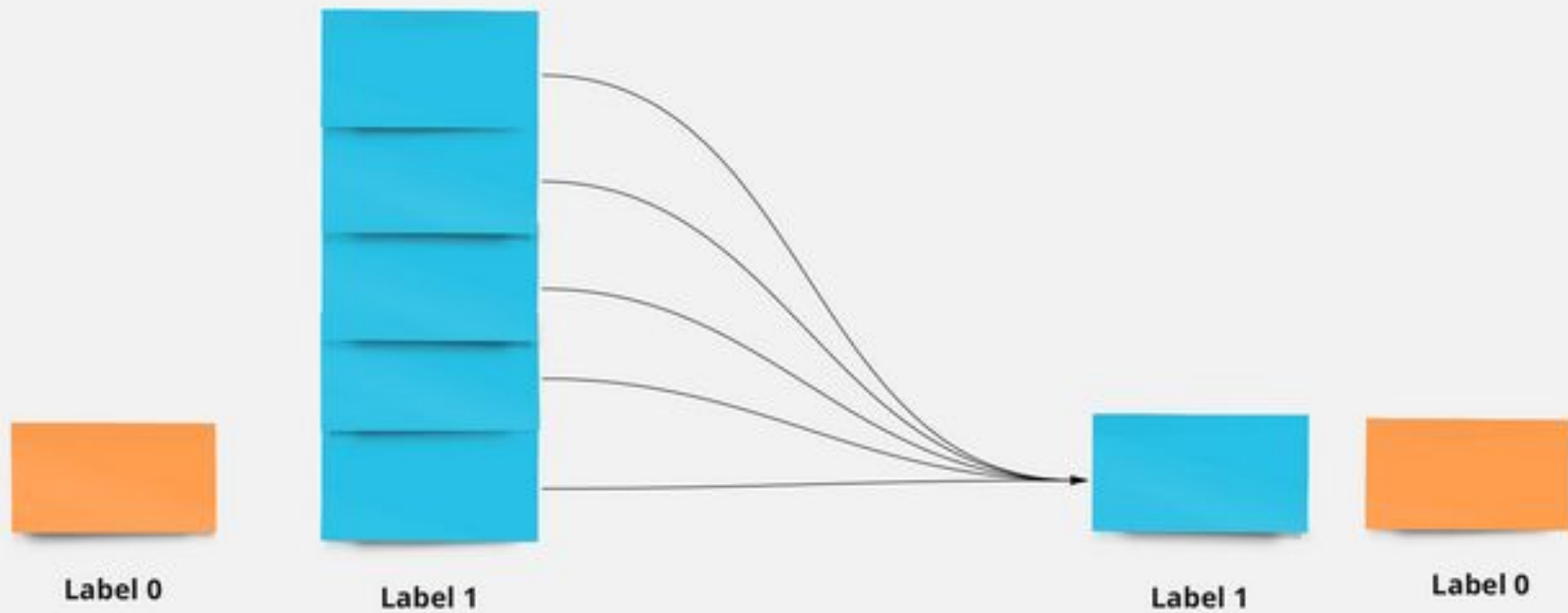
No
fraudulentas

Fraudulentas

599224

776

Undersampling

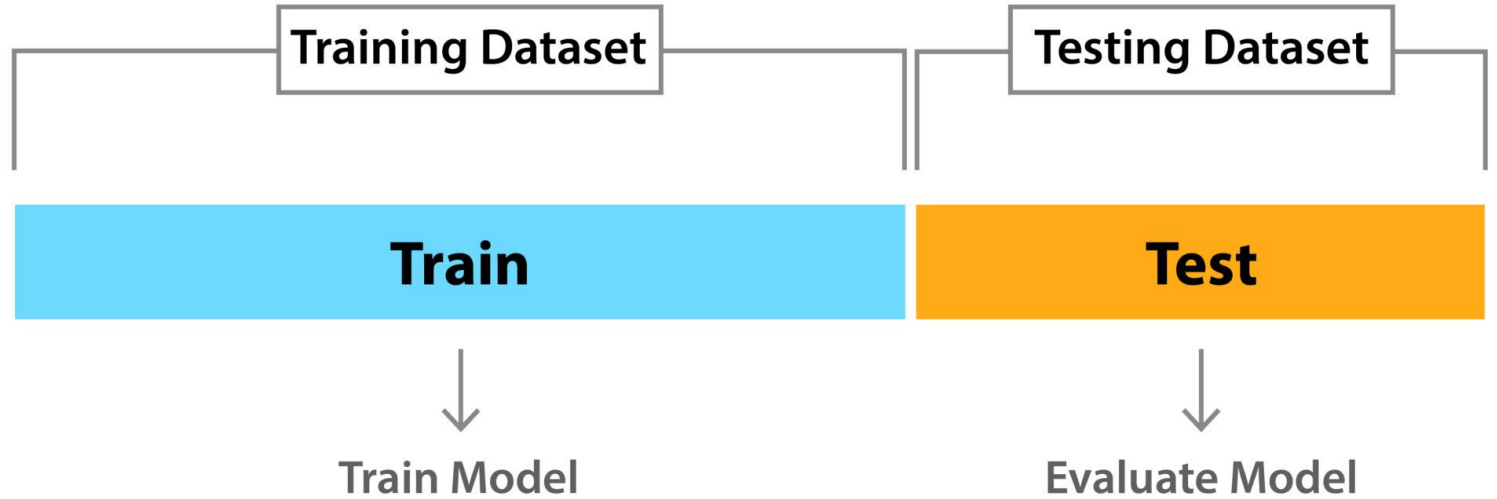




División de datos

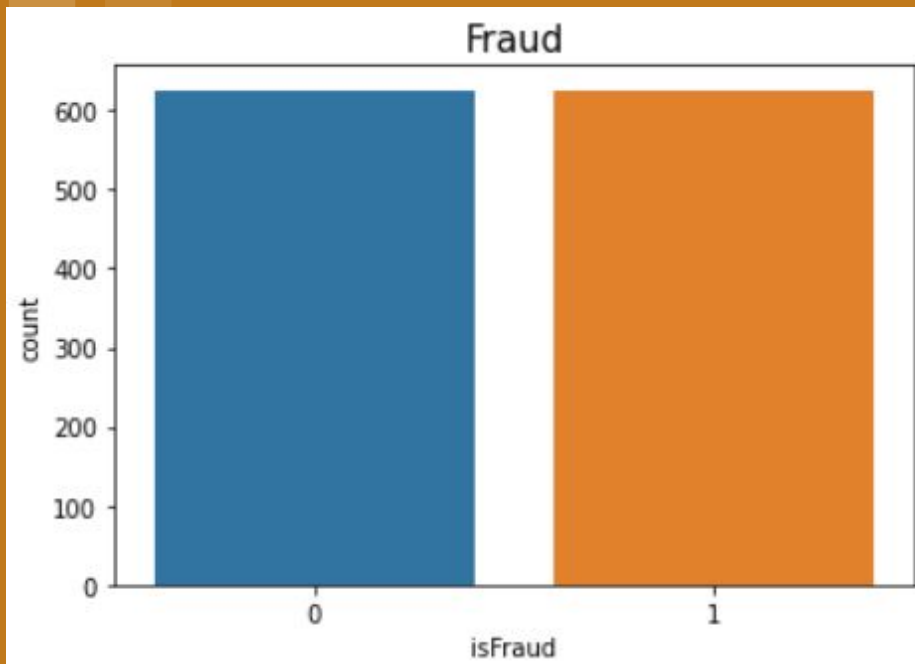


DATASET



Train	Test
80%	20%

Train



626

Test



150



Valores nulos

Type	isFraud
0	0



Variables innecesarias





Variable	Porcentaje de valores únicos en el dataset
nameOrig	99.99%
nameDest	72.86%

Ordinal Encoding

Variable “type”	Variable “type”
Payment	0
Transfer	1
Cash in	2
Cash Out	3
Debit	4

One Hot Encoding

type_CASH_IN	type_CASH_OUT	type_DEBIT	type_PAYMENT	type_TRANSFER
1	0	0	0	0
0	0	0	1	0
0	1	0	0	0
0	0	0	0	1
1	0	0	0	0



Resultados

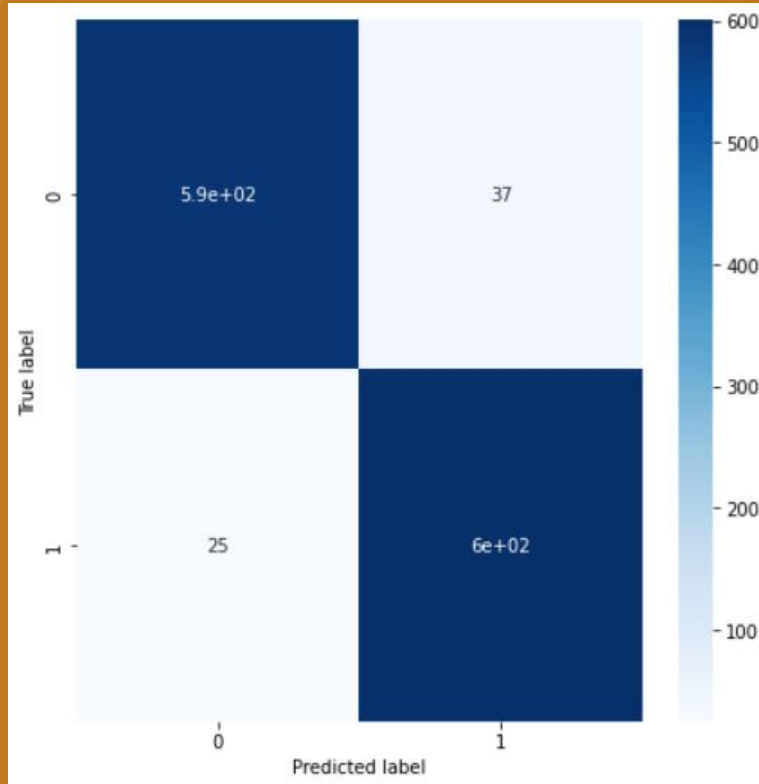




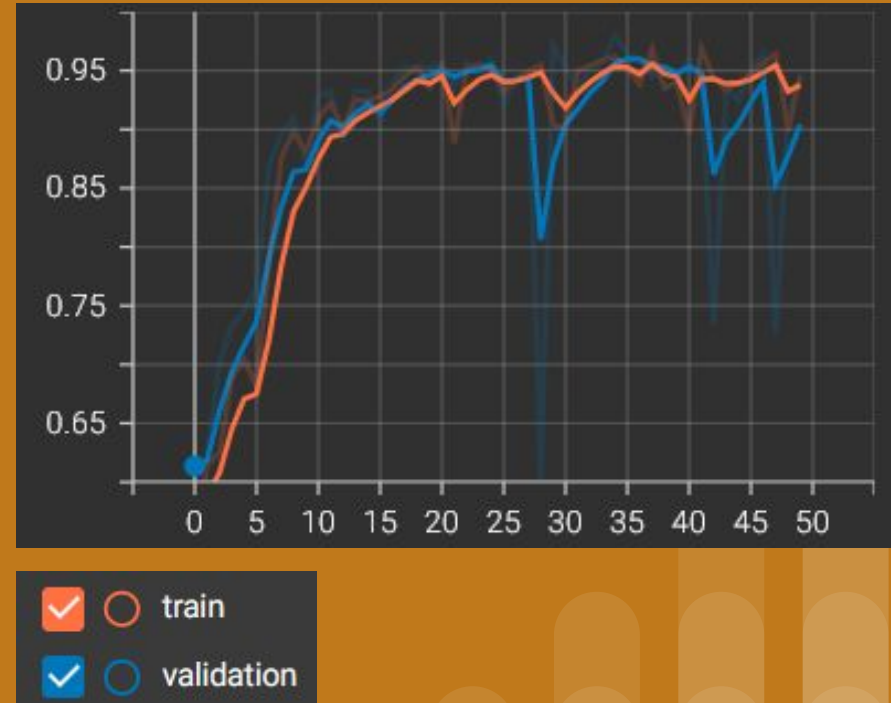
Ordinal Encoding y Undersampling



Accuracy: 94%



Accuracy por Epoch



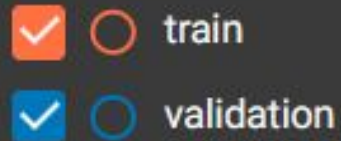
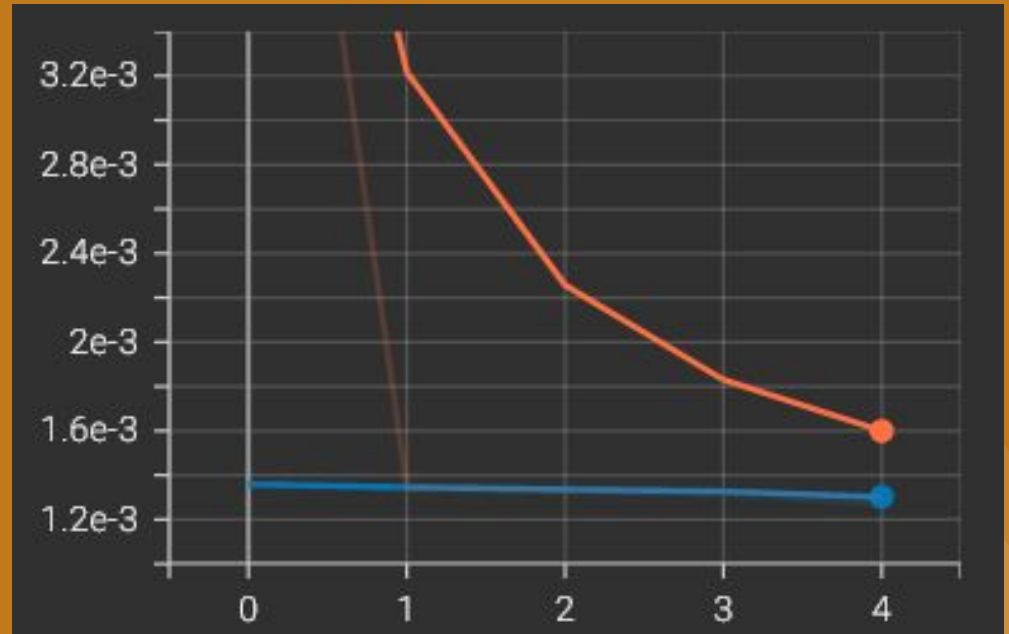


Ordinal Encoding y Autoencoders




Accuracy: 91%

Loss por Epoch

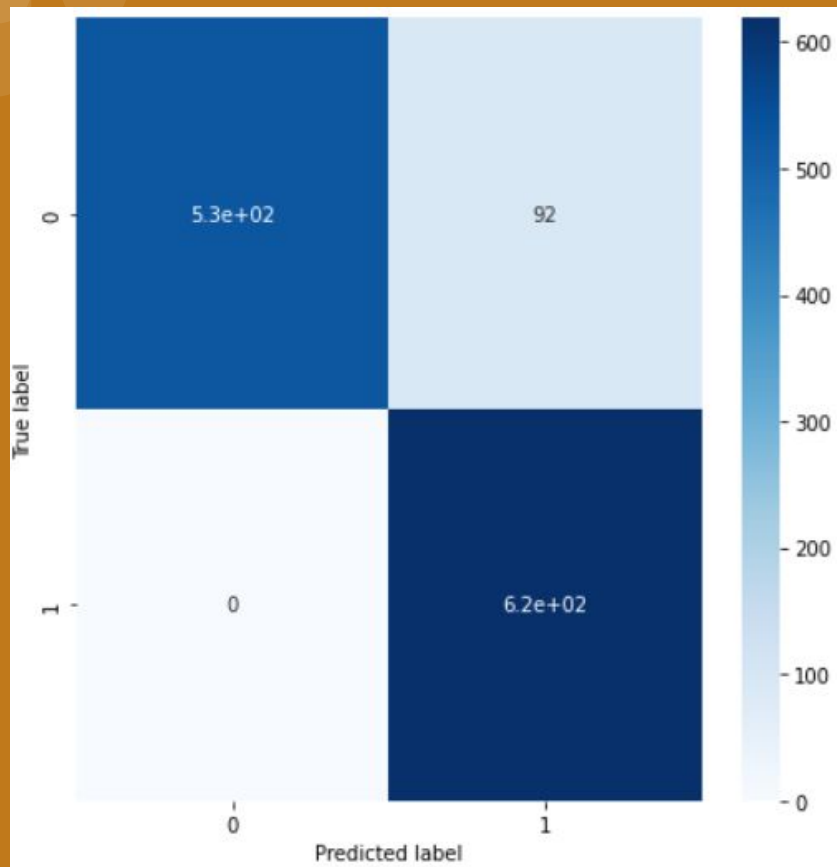




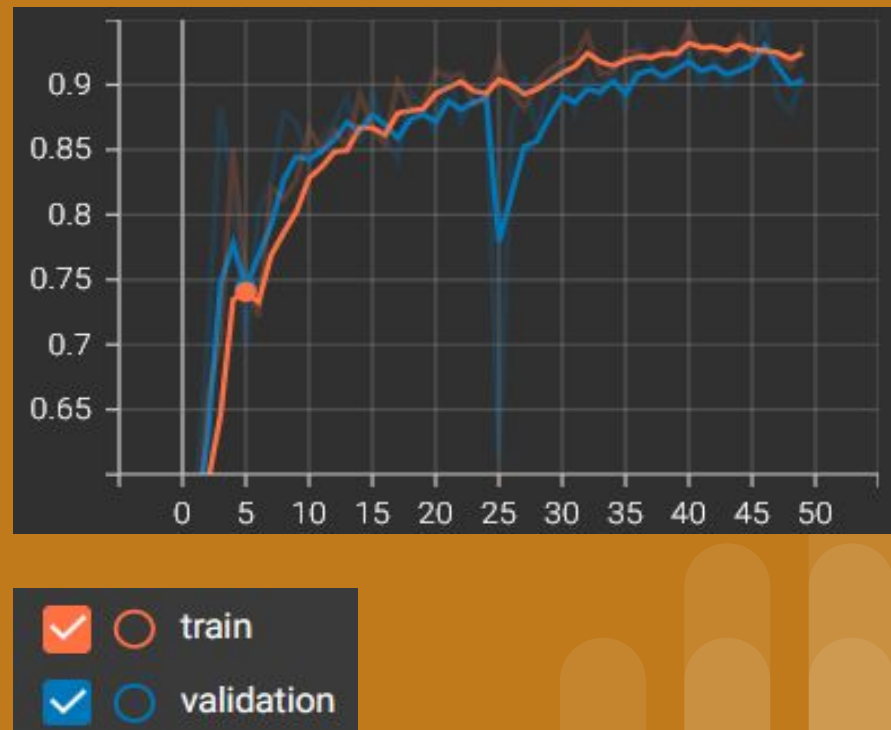
One Hot Encoding y Undersampling



Accuracy: 91%



Accuracy por Epoch



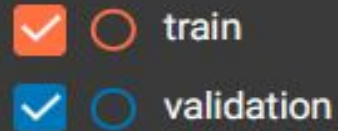
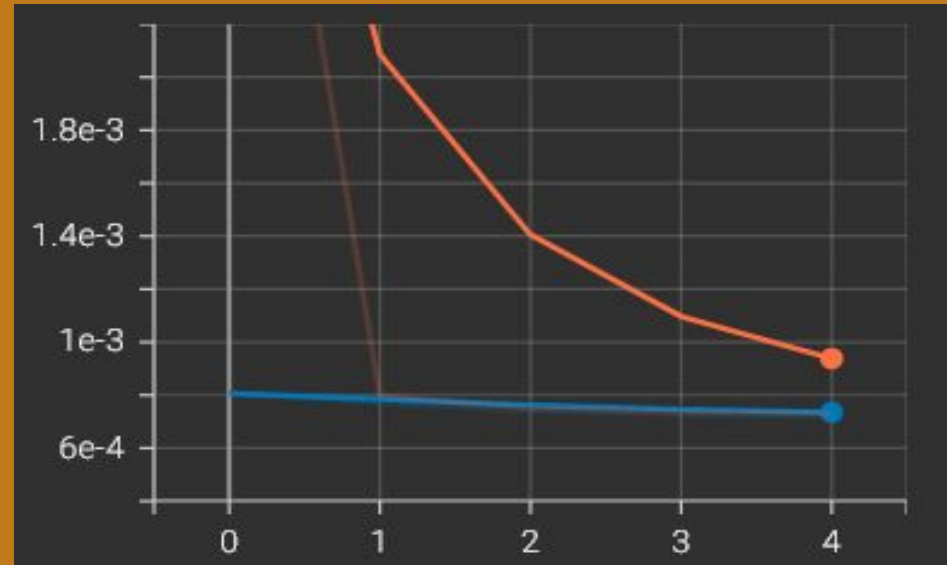


One Hot Encoding y Autoencoders



Accuracy: 92%

Loss por Epoch



Modelos y Accuracies

	Undersampling	AutoEncoders
Ordinal Encoder	94%	91%
One Hot Encoder	91%	92%

Conclusiones

1. El número de reportes de transacciones fraudulentas va en aumento dia con dia. Por lo tanto, es necesaria una solución eficiente y eficaz.
2. El mejor modelo para clasificar y/o detectar que una transacción es fraudulenta es utilizando Ordinal Encoding y Undersampling en los datos.
3. Las redes neuronales son herramientas que ayudan a detectar una infinidad de anomalías en un conjunto de datos, por lo que son muy utilizados en la solución a distintos problemas, de esta forma reducir el impacto de estos.