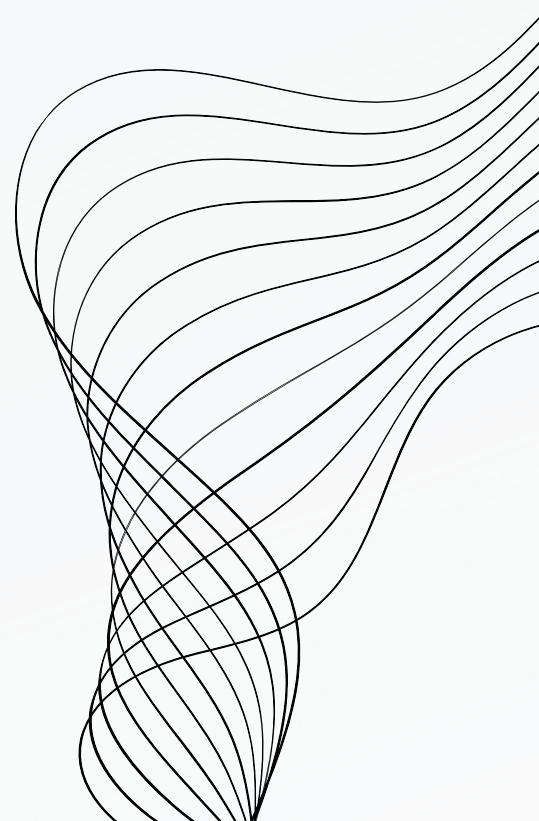




**PWC CONSULTING CHALLENGE**

# **DETECCIÓN DE FRAUDE EN TARJETAS DE CRÉDITO**

**NICOLÁS BAZTÁN YOLDI  
VÍCTOR MONTILLA CASTILLA  
ALEJANDRO PALANCAR DEL ESTAL**

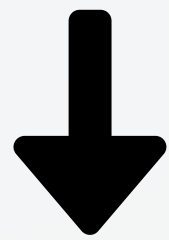


# ÍNDICE

- 01** INTRODUCCIÓN Y OBJETIVO
- 02** ADQUISICIÓN DE DATOS
- 03** ANÁLISIS EXPLORATORIO DE DATOS (EDA)
- 04** INGENIERÍA DE CARACTERÍSTICAS
- 05** PREPARACIÓN DE DATOS PARA MODELADO
- 06** SELECCIÓN Y ENTRENAMIENTO DEL MODELO
- 07** EVALUACIÓN DEL MODELO
- 08** VALIDACIÓN Y OPTIMIZACIÓN DEL MODELO
- 09** CONCLUSIONES Y CONSIDERACIONES FINALES

# INTRODUCCIÓN Y OBJETIVO

- Transacciones digitales desplazando gradualmente al dinero en efectivo.
- Discernir movimientos financieros es crucial.



- **Reto:** detección de fraude en tarjetas de crédito.



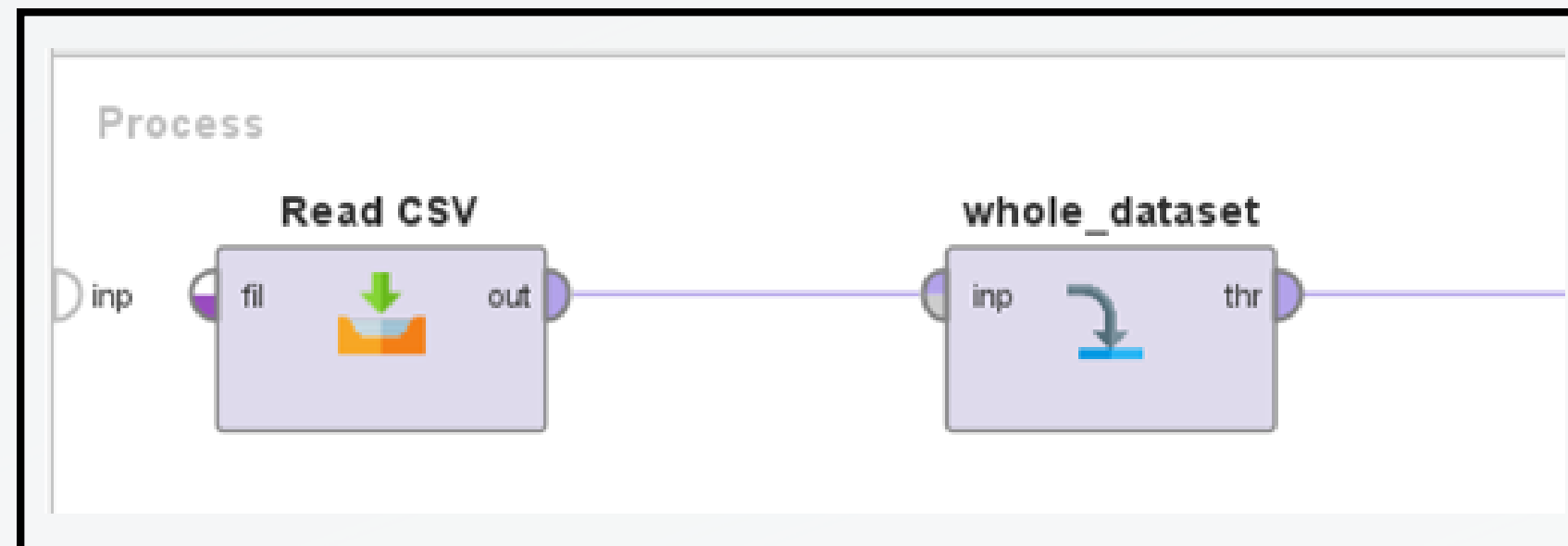
*Técnicas de Machine Learning*



- RapidMiner

# ADQUISICIÓN DE DATOS

## Obtención de datos



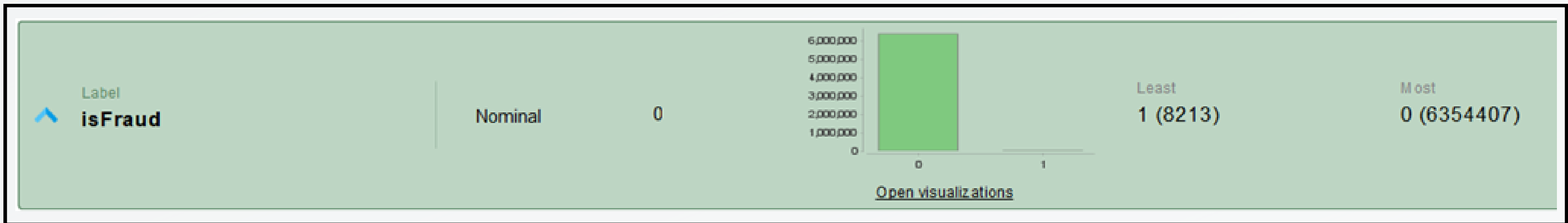
- a) Nombre: whole\_dataset.csv
- b) Tipo: CSV
- c) Observaciones: 6362620
- d) Atributos: 11 atributos

## Descripción del DataSet

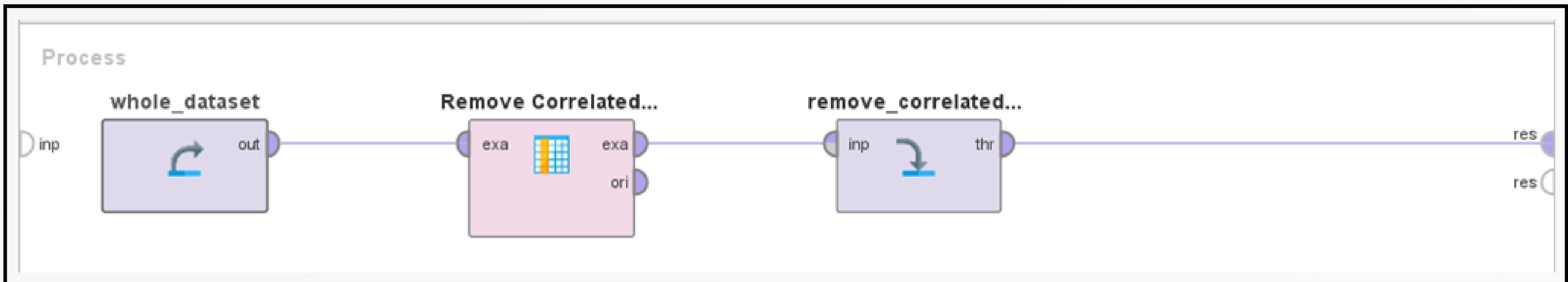
Name	Type	Missing
amount	Real	0
oldbalanceOrg	Real	0
newbalanceOrig	Real	0
oldbalanceDest	Real	0
newbalanceDest	Real	0
step	Integer	0
isFlaggedFraud	Integer	0
isFraud	Nominal	0
type	Nominal	0
nameOrig	Nominal	0
nameDest	Nominal	0

# ANÁLISIS EXPLORATORIO DE DATOS (EDA)

## Descripción de la variable 'isFraud'

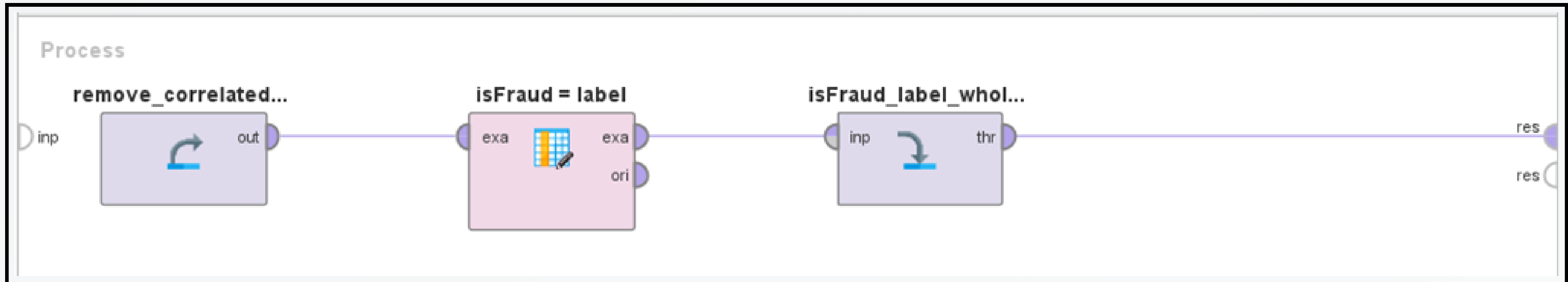


## Eliminación de correlaciones entre variables



# INGENIERÍA DE CARACTERÍSTICAS

*Eliminación de correlaciones entre variables*



- Ajuste de la muestra de datos para futuros modelos.
- 'isFraud' = punto central alrededor del cual girarán los modelos.





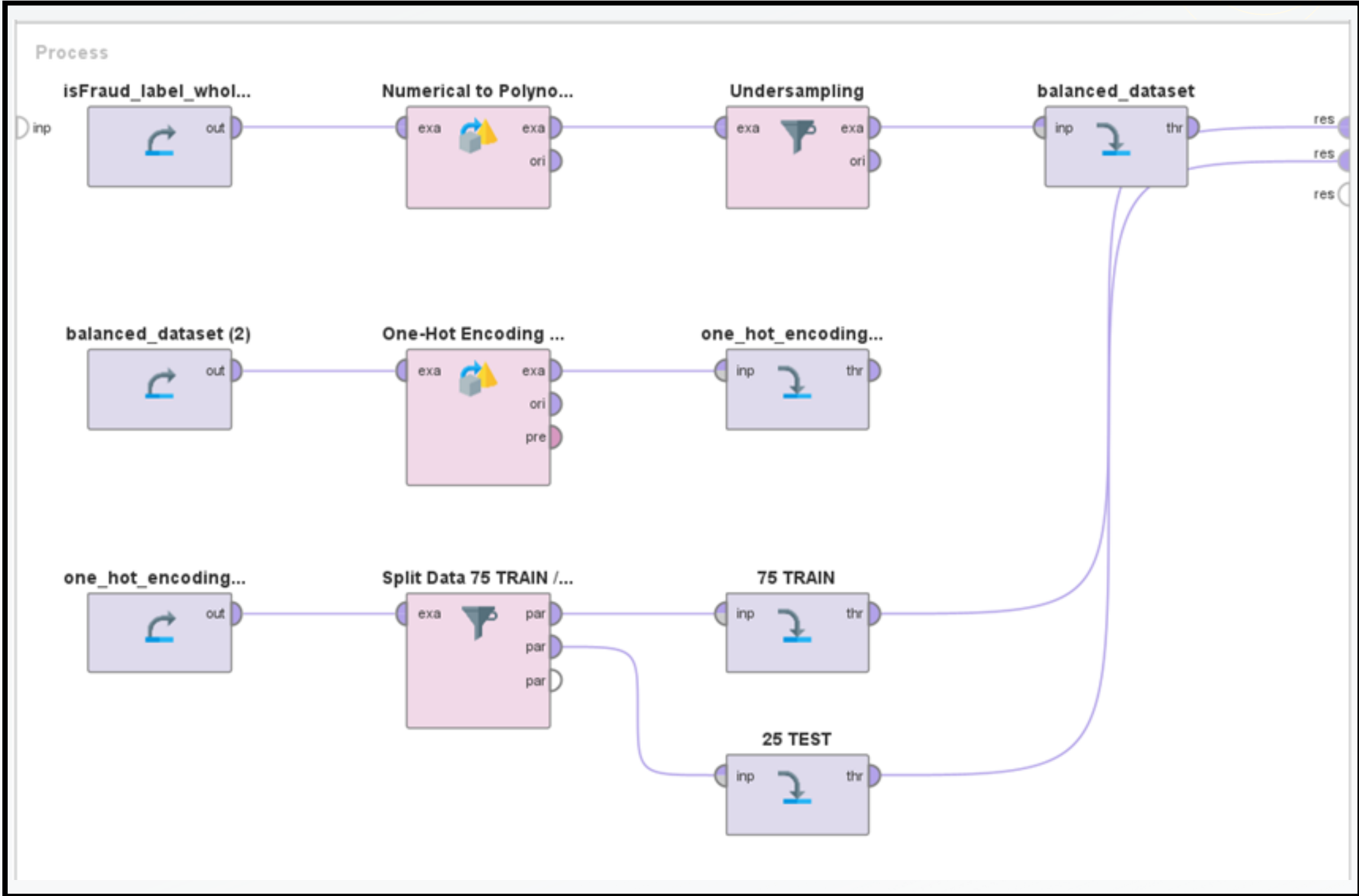
# PREPARACIÓN DE DATOS PARA MODELADO

## Undersampling – Balanceo del DataSet

## Codificación de variables categóricas

## *División del DataSet balanceado*

- Muestra de entrenamiento (75%)
- Muestra de validación (25%)



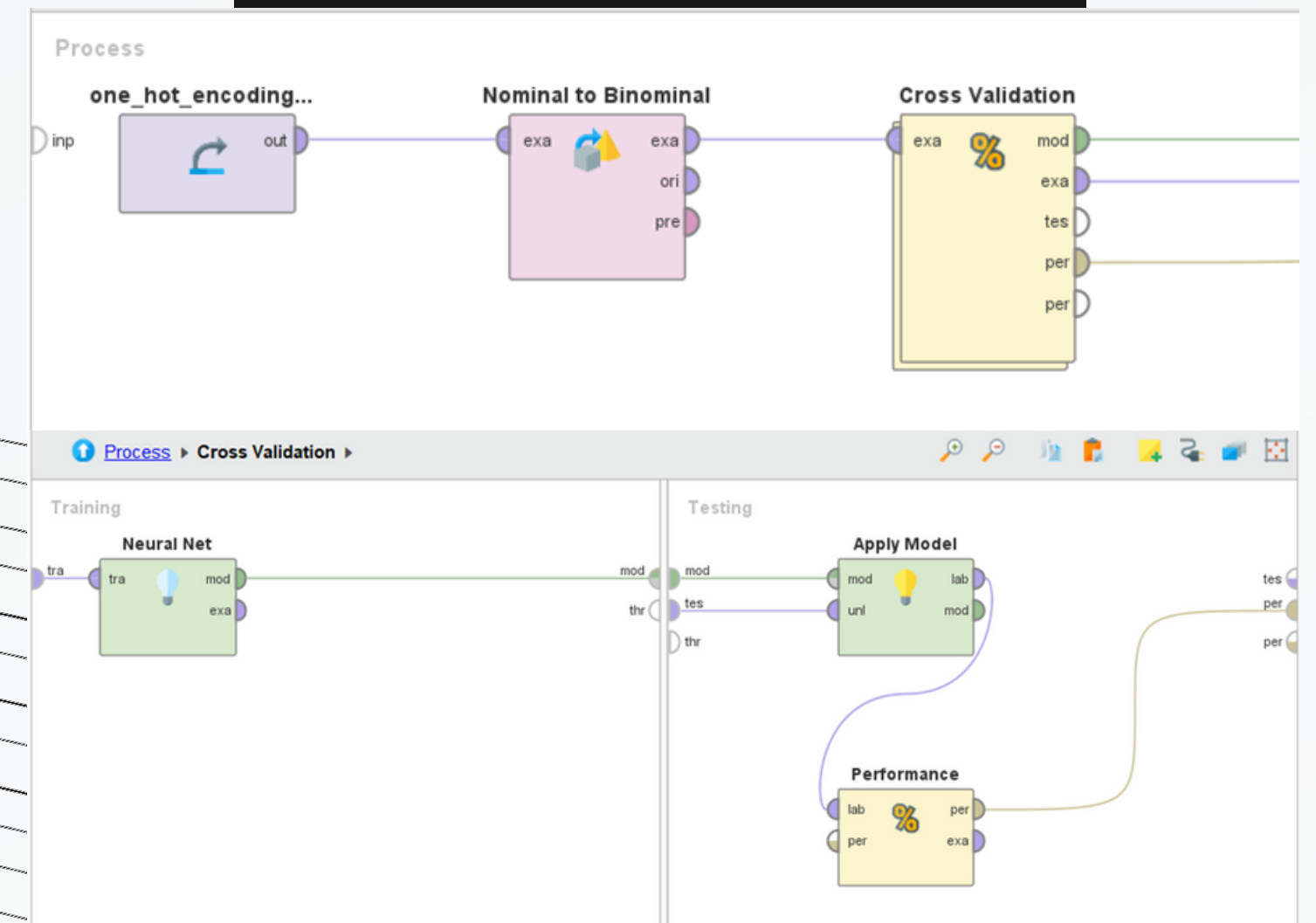
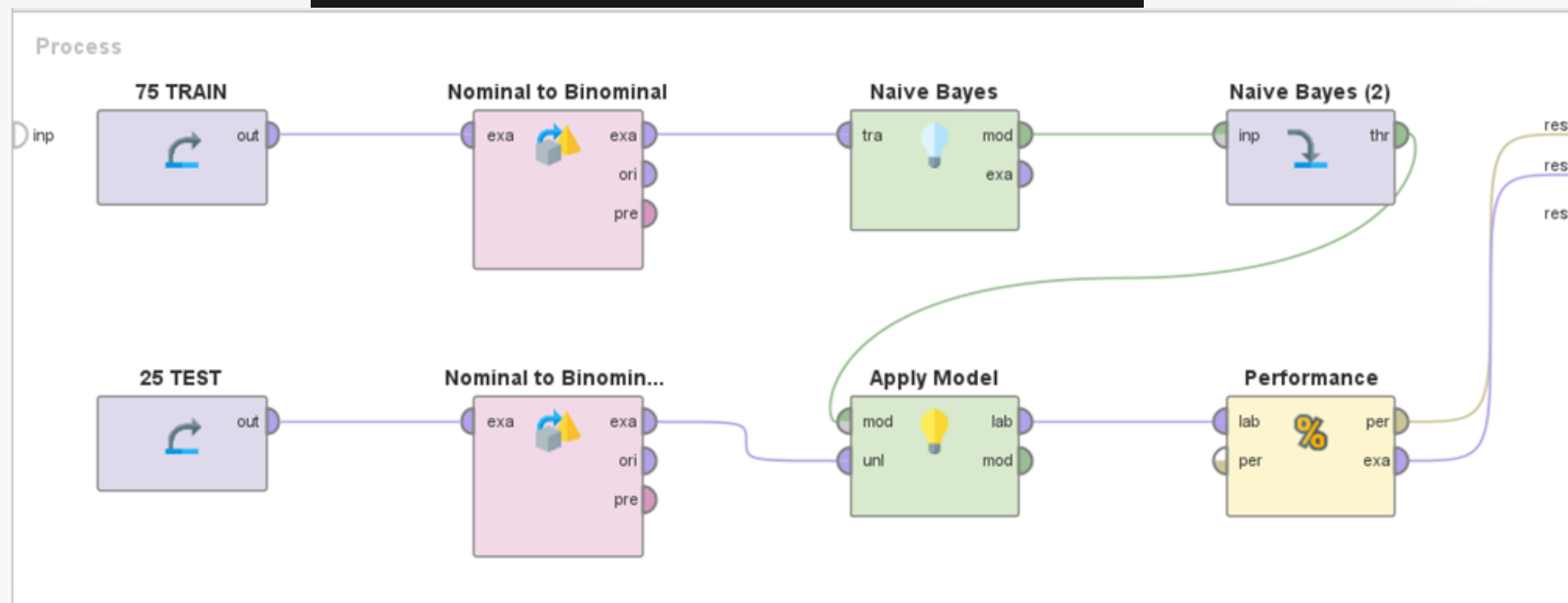
# SELECCIÓN Y ENTRENAMIENTO DEL MODELO

## NAIVE BAYES

- Fácil comprensión y gran rendimiento
- Asume independencia
- TRAIN 75%
- TEST 25%

## NEURAL NET

- Complejo y profundo
- Cross Validation



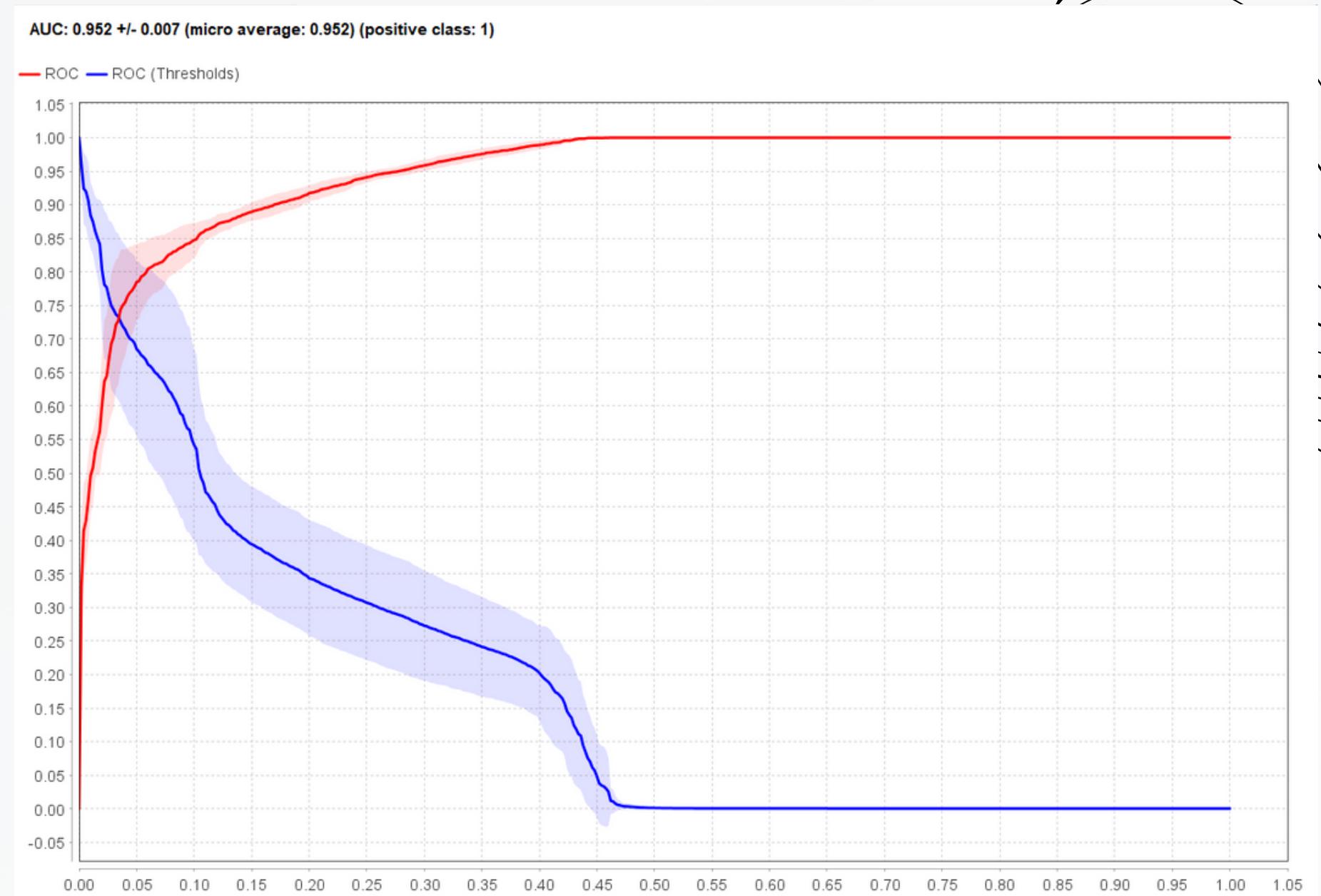
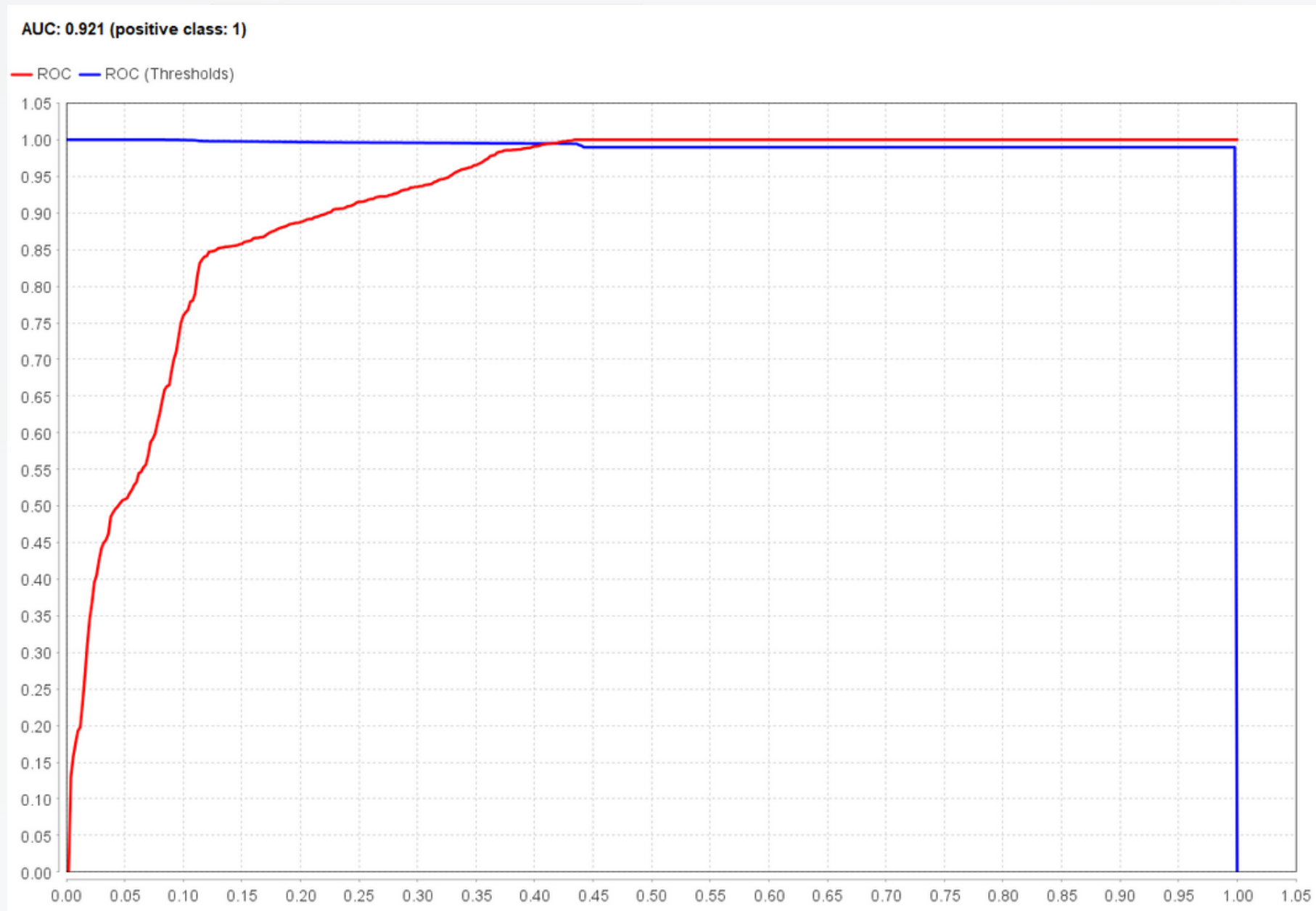


# EVALUACION DEL MODELO

NAIVE BAYES (92,1%)

AUC

NEURAL NET (95,2%)



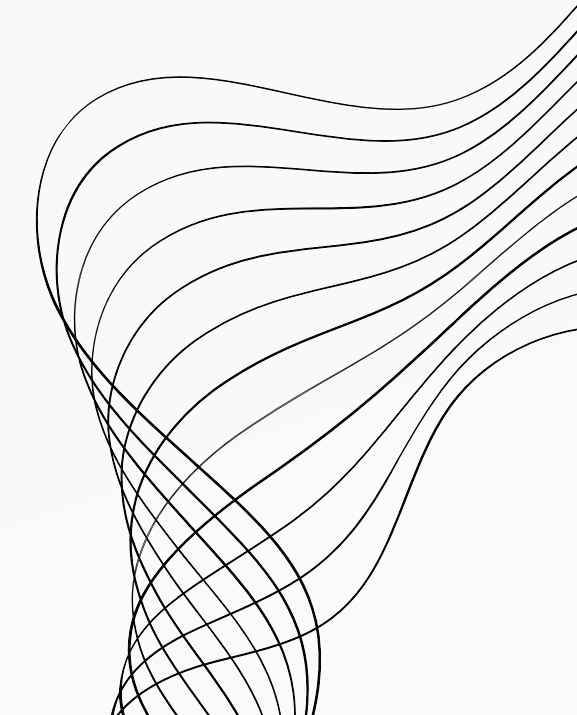
# VALIDACIÓN Y OPTIMIZACIÓN DEL MODELO

		LEARNING RATE				
MOMENTUM	AUC (%)	0.1	0.2	0.3	0.4	0.5
	0.1	95,2	95,8	95,7	95,8	95,7
	0.2	95,4	95,8	95,9	95,7	95,7
	0.3	95,5	95,8	95,8	95,7	95,6
	0.4	95,5	95,8	95,7	95,8	95,2
	0.5	95,7	95,8	95,7	95,3	95,3

TRAINING CYCLE 500	97,7
TRAINING CYCLE 1000	98,4
TRAINING CYCLE 1500	98,7
TRAINING CYCLE 2000	98,8



# **CONCLUSIONES:**

- **PREDICCIÓN DE FRAUDES EN TARJETAS DE CRÉDITO**
  - **TÉCNICAS DE MACHINE LEARNING PARA MITIGAR RIESGOS FINANCIEROS**
  - **MODELOS EFECTIVOS PARA LA DETECCIÓN DE FRAUDES**
- 



**PWC CONSULTING CHALLENGE**

# **DETECCIÓN DE FRAUDE EN TARJETAS DE CRÉDITO**

**NICOLÁS BAZTÁN YOLDI  
VÍCTOR MONTILLA CASTILLA  
ALEJANDRO PALANCAR DEL ESTAL**

