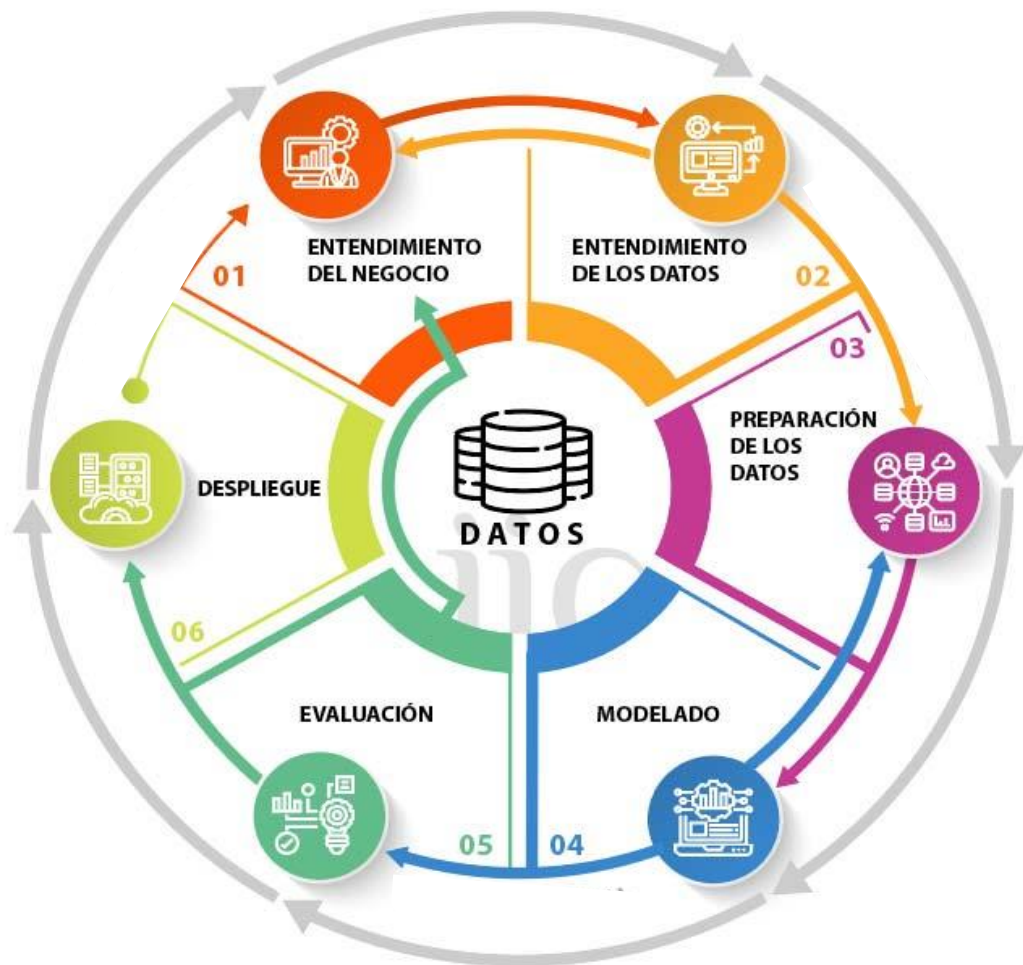




# Modelo Detección fraccionamiento transaccional



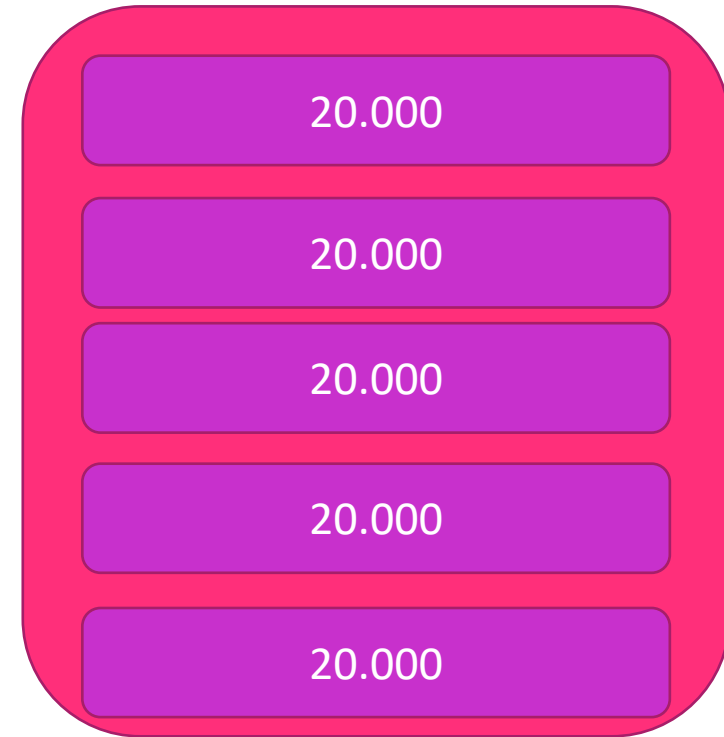
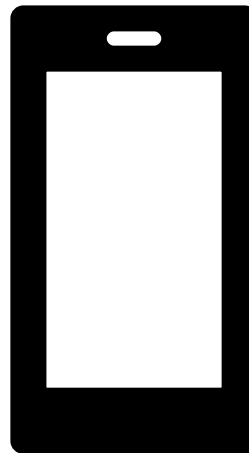
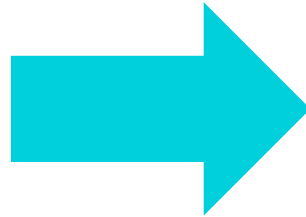
# ¿Qué estamos buscando?

## Detectar el fraccionamiento transaccional

# Fraccionamiento Transaccional



1 Transacción



5 Transacciones

## Condiciones

- Mismo corresponsal
- Misma cuenta
- Menos de 24 horas

# Entendimiento de los datos

## 3. Variables

- `_id`: Identificador único del registro
- `merchant_id`: Código único del Comercio o aliado.
- `subsidiary`: Código unico de la sede o sucursal.
- `transaction_date`: Fecha de contabilización de la transacción en el core Financiero.
- `account_number`: numero único de cuenta
- `user_id`: Código único del usuario dueño de la cuenta desde donde se registran las transacciones.
- `transaction_amount`: Monto de la transaction en moneda ficticia.
- `transaction_type`: Naturaleza de la transacción valores (DEBITO o CREDITO)

# Preparación de los datos

Para identificar la mala práctica transaccional de los corresponsales, se hace un agrupamiento que va contar por comercio y número de cuenta las transacciones que se realizaron en un lapso de 24 horas. Se hace un conteo para saber que cantidad se hicieron.

## Condición

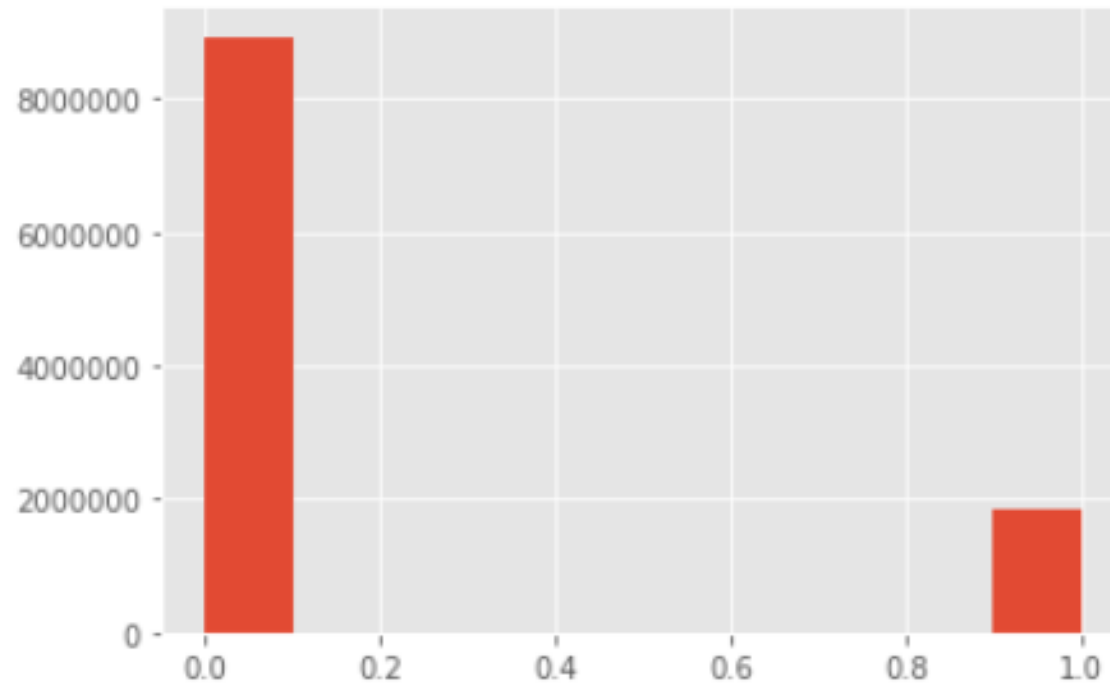
Si se hicieron mas de dos transacciones por el mismo corresponsal, a una misma cuenta durante las próximas 24 horas, se clasifica como fraude la transacción.



# Variable respuesta

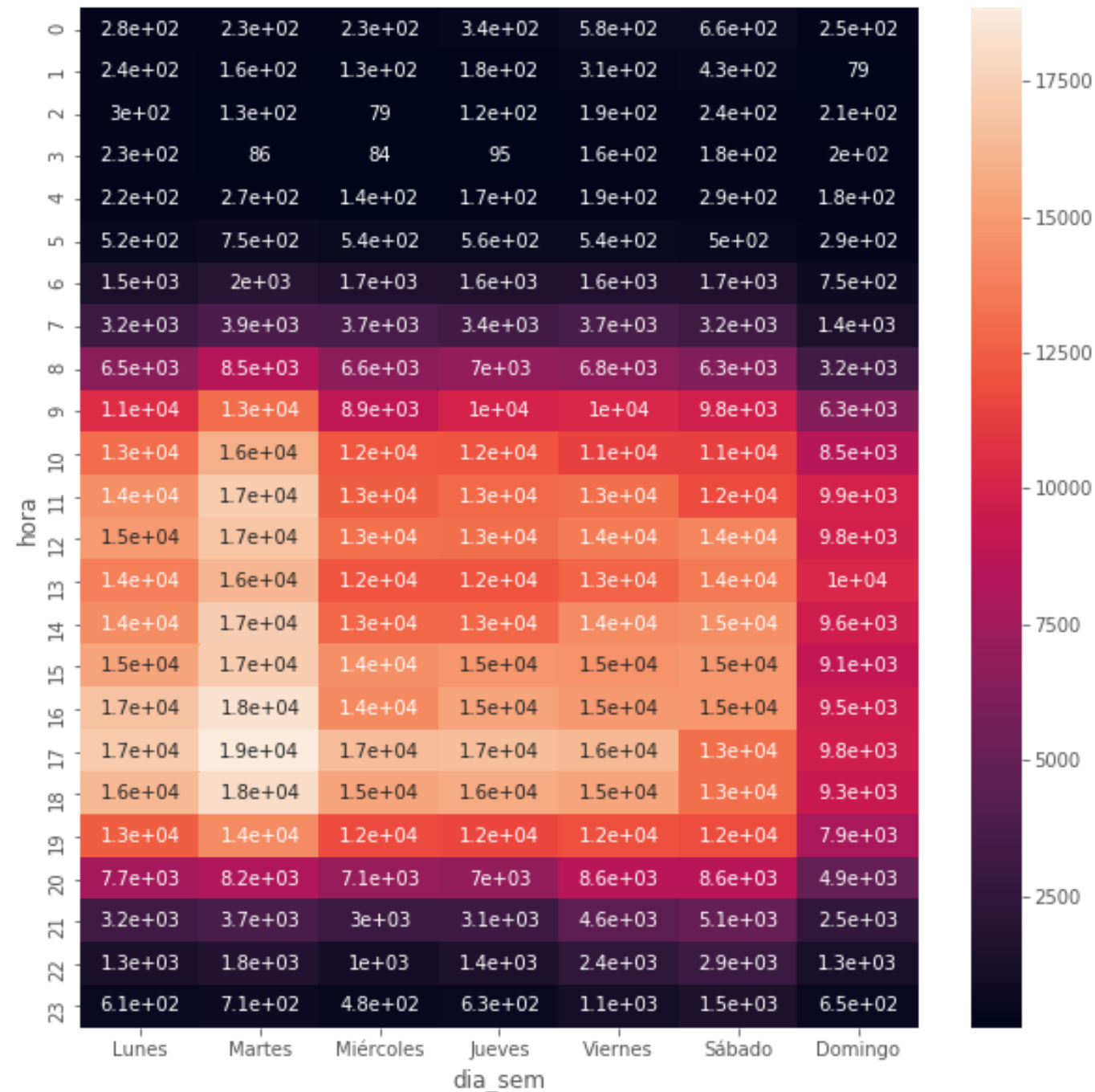
0: No fraude **83%**

1: Fraude **17%**



**No Fraude**

**Fraude**



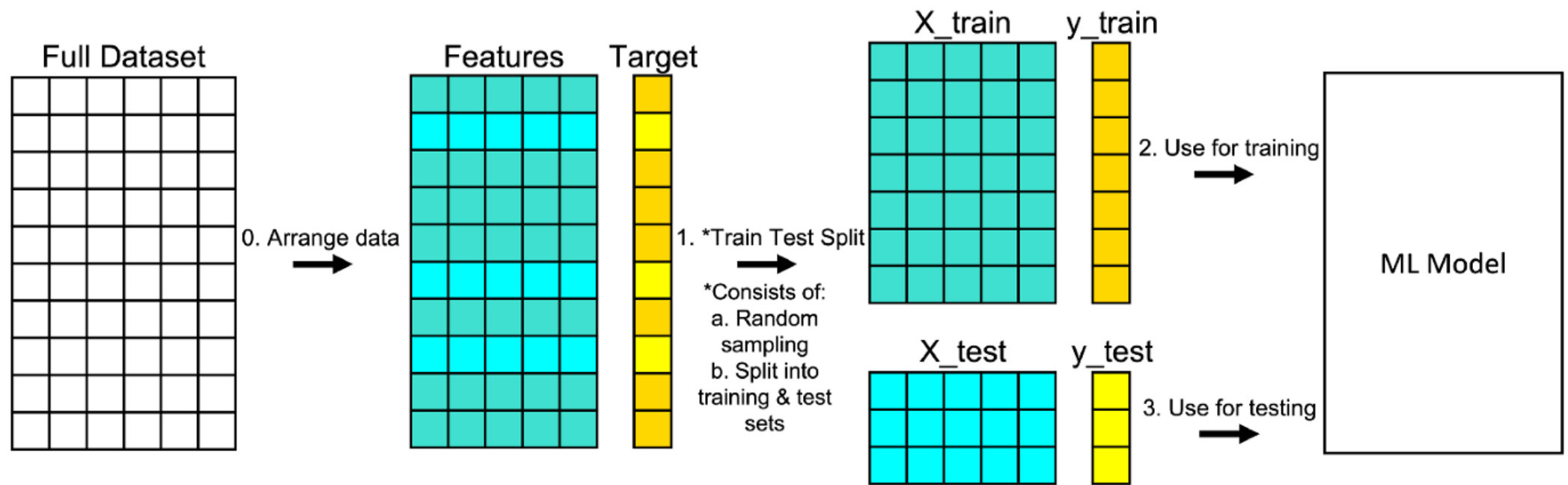
**Hora y día de la semana donde mayor cantidad de transacciones se han hecho.**

**Posible mayor fraccionamiento**

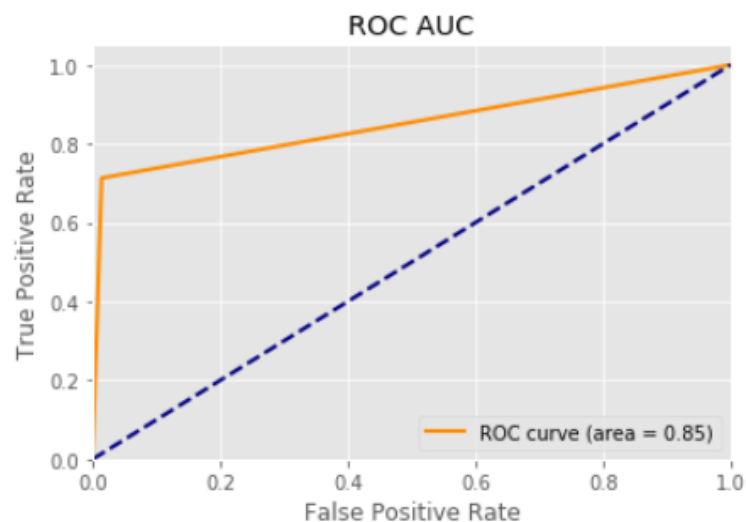
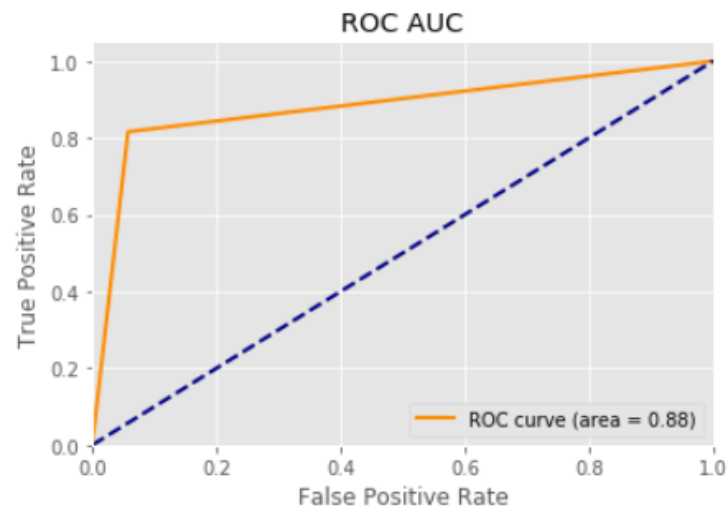
# Modelado



# Metodología



# Selección Modelo

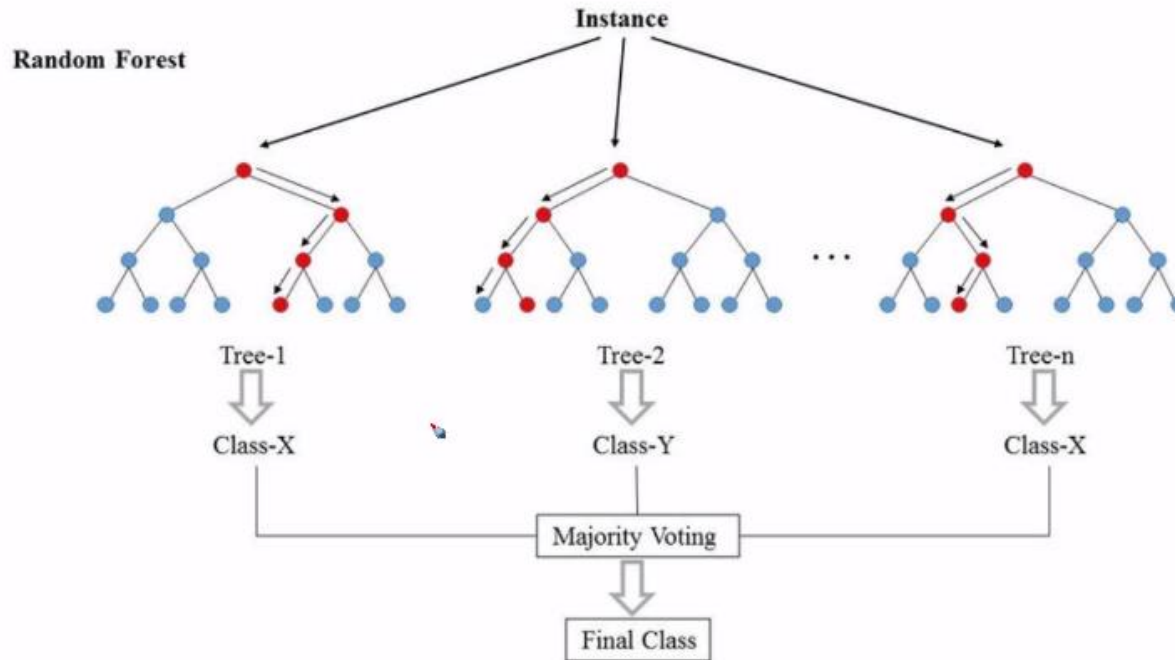


	F1	ROC AUC
Random Forest con class_weight	0.7777	0.8791
Xgboost	0.8004	0.8494
Random Forest	0.7986	0.8492
Decision Tree	0.7972	0.8482
Regresión logística con Class weighting	0.7371	0.8402
Ada Boost	0.7902	0.8371
Gradient Bosting	0.7849	0.8265
Regresión logística select variables	0.7736	0.8154
Xgboost con hiperparámetros	0.7736	0.8154

# Random Forest

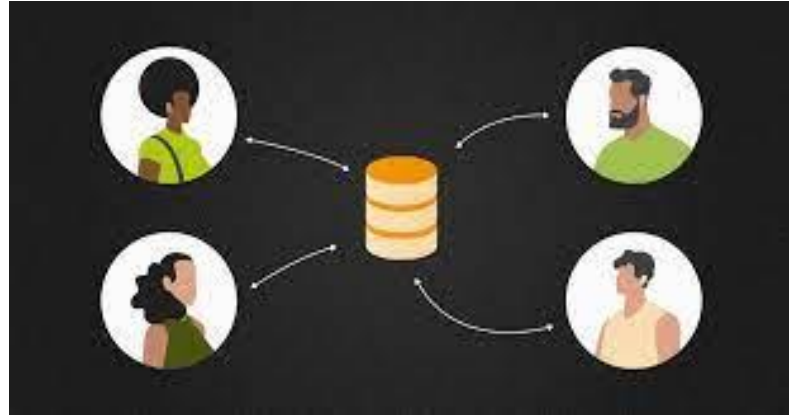
## (Bosque Aleatorio)

Los Random Forests tienen una **capacidad de generalización muy alta** para muchos problemas.



- ✓ Es un conjunto (ensemble) de árboles de decisión combinados.
- ✓ Distintos árboles ven distintas porciones de los datos.
- ✓ Ningún árbol ve todos los datos de entrenamiento. Esto hace que cada árbol se entrene con distintas muestras de datos para un mismo problema.
- ✓ Al combinar sus resultados, unos errores se compensan con otros y tenemos una predicción que generaliza mejor.

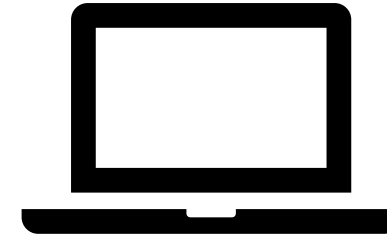
# Arquitectura



**Recolección de datos**

**24 horas**

**API**



**Modelo  
PIPELINE**

**hubo  
fraccionamiento**

**No hubo  
fraccionamiento**

**Reentrenamiento  
nuevos datos**

**Ingesta de  
info**