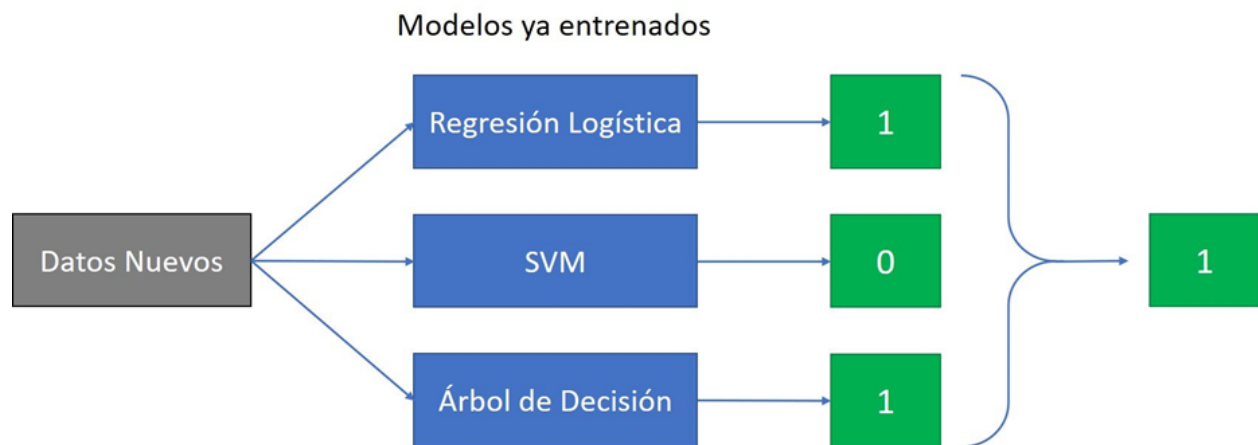


# Ensamblas Híbridos

- **Ensamblas homogéneos:** combinan el mismo tipo de modelo
  - Bagging
  - Boosting
- **Ensamblas híbridos:** combinan clasificadores de distinto tipo
  - Voting
  - Stacking
  - Cascading

## Voting

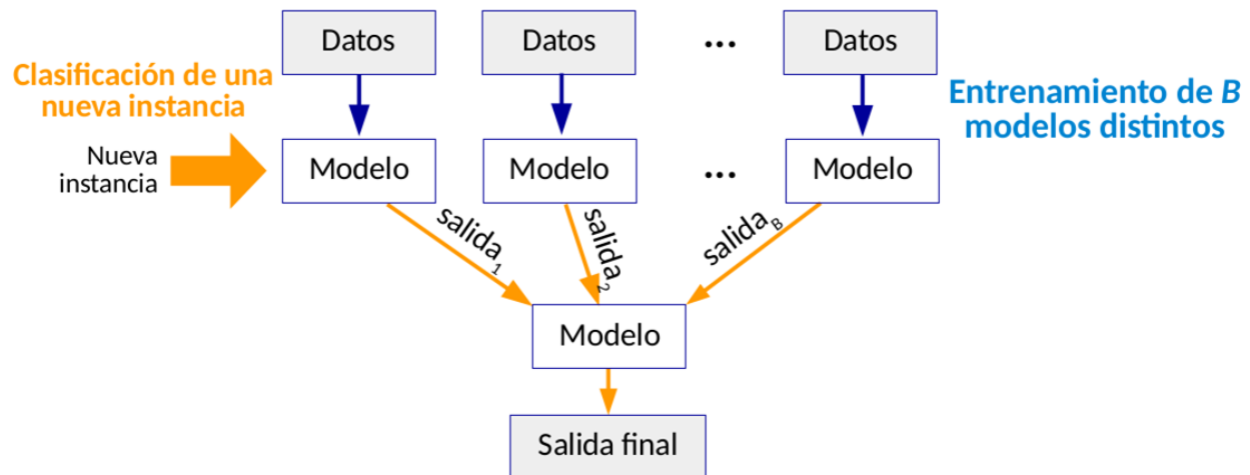
Construir N modelos utilizando los mismos datos y luego tomar la predicción mayoritaria



## Stacking

Entrenar diferentes modelos (modelos base) y un modelo más que decide, dada una instancia nueva, qué modelo usar.

- Concepto de meta-aprendizaje para reemplazar el mecanismo de voto
- Pueden apilarse tantas capas de modelos como se desee



- Los modelos para el meta-aprendizaje suelen ser: árboles, NB, SVM o Perceptrón (en Redes Neuronales)
  - Para los otros puede usarse cualquiera
  - Menos popular que boosting, bagging
    - Dificultad de análisis teórico: caja negra
    - Múltiples variantes
  - Se puede interpretar como una mejora (generalización) del método de votación (Voting)
  - Si los clasificadores base pueden generar medidas de certeza, suele funcionar mejor
- A veces en los modelos podemos decirles que no solo nos de la categoría, sino con que probabilidad cree el modelo que es de esa categoría.

# Cascading

- Enfoque en el que se pasa sucesivamente los datos de un modelo a otro
  - A diferencia de Stacking cada “capa” tiene un sólo modelo
  - El input de cada modelo son las instancias predichas con poca certeza por el modelo anterior
- 
- Suele utilizarse cuando se necesita una alta certeza en la predicción

## Ejemplo

Queremos un modelo que prediga si una transacción con tarjeta de crédito es fraudulenta

Necesitamos una alta certeza para definir si no lo es, en caso de error las pérdidas pueden ser millonarias

Lo mismo sería con un modelo donde queremos detectar casos de pacientes con cancer.

No sería tan grave predecir como positivo un caso negativo, porque después si lo ve un experto lo va a descartar → es un falso positivo

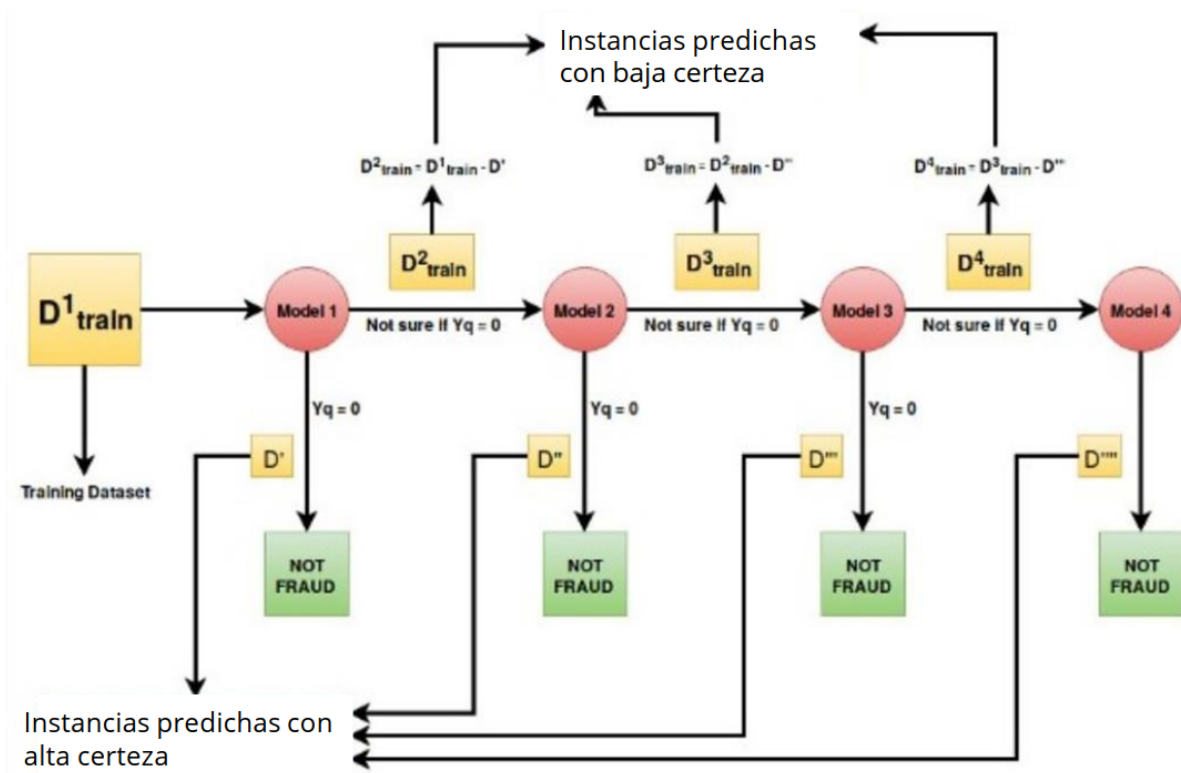
Pero si es grave descartar uno que era un caso positivo, este error puede costar una vida.

Construimos entonces una secuencia (o cascada) de modelos de ML

## Entrenamiento

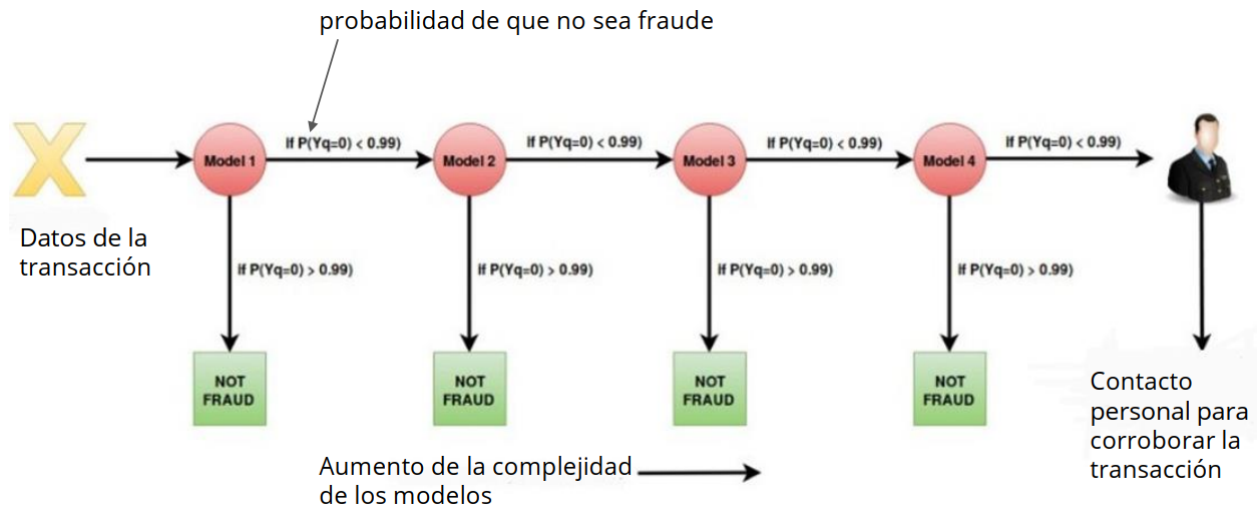
- Entrenamos N modelos distintos utilizando nuestro set de entrenamiento
- Cada modelo lo entrenamos sobre las instancias predichas con baja certeza del modelo anterior

- Suele hacerse de forma tal de empezar por modelos simples y a medida que se entrenan nuevos los mismos sean de mayor complejidad



## Predicción

- El primer modelo recibe los datos de la transacción
- Si la probabilidad de que no sea fraude es menor a 0.99, se pasa al siguiente modelo
- Caso contrario se descarta la posibilidad de fraude
- Si ningún modelo descarta el fraude con  $p < 0.99$  se procede a confirmar la operación de forma personal



- A diferencia de Voting y Stacking que tienen un enfoque de modelos “Multi-expertos”, Cascading tiene un enfoque “Multi-estado”
- Inicialmente creados para computer vision
- Cascadas muy profundas pueden producir overfitting

## Resumen

- La combinación de clasificadores permite generar clasificadores más precisos a cambio de menor comprensión de los mismos
- Métodos homogéneos básicos: bagging, boosting
  - Combinan mismo tipo de clasificador
  - Manipulación del conjunto de entrenamiento
- Método heterogéneo básico: Voting
  - Combinan distintos tipos de clasificador

## Propiedades de los Ensamblados

- Ventajas

- Generalmente logran mejores predicciones que los modelos vistos hasta ahora (random forest, xgboost)
- Buen trade-off sesgo varianza

- Desventajas

- Se pierde interpretabilidad
- Tienen una complejidad computacional mayor

### Ensembles: voting, bagging, boosting, stacking - IArtificial.net

Un ensemble es un conjunto de modelos de machine learning. Cada modelo produce una predicción diferente. Las predicciones de los distintos modelos se combinan para obtener una única predicción. La

IA <https://www.iartificial.net/ensembles-voting-bagging-boosting-stacking/>

