

# DATA SCIENCE



---

Federico Baiocco  
baioccofed@gmail.com  
3512075440



# Clase 19 - Agenda

MÉTRICAS DE EVALUACIÓN PARA  
CLASIFICACIÓN

---

# ¿ Dudas de la clase pasada ?

---

¿ Todos pudieron terminar ?



# Clasificación binaria



En problemas de clasificación binaria, buscamos separar los elementos de un conjunto de datos en 2 clases. Por ejemplo, lo que estuvimos trabajando en clases anteriores de titanic donde buscábamos clasificar a los pasajeros en sobrevivieron / no sobrevivieron.

Por ejemplo:

- Exámenes: Aprobado / no aprobado
- Exámenes médicos: Enfermo / Sano
- Etc

# Clasificación binaria



¿ Qué resultados podemos obtener en un problema de clasificación binaria ?  
(Por ejemplo, en un test de embarazo):

- **Verdadero** positivo: Test positivo, persona embarazada
- **Verdadero** negativo: Test negativo, persona no embarazada
- **Falso** positivo: Test positivo, persona no embarazada
- **Falso** negativo: Test negativo, persona embarazada

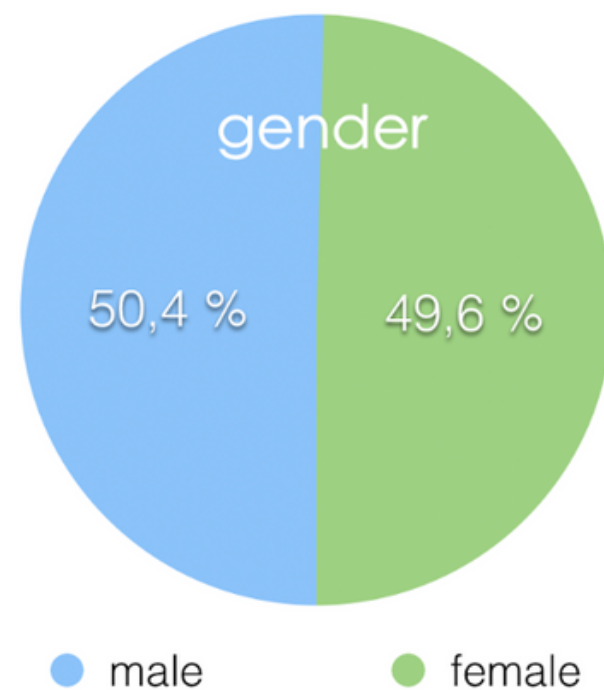
# Limitaciones de accuracy score



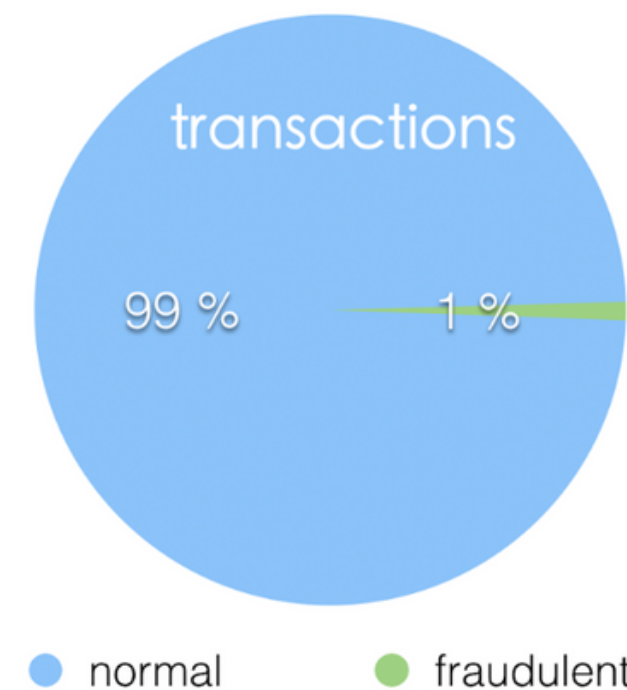
¿ Se imaginan qué limitaciones tiene el accuracy score que estuvimos usando hasta ahora ?

PISTA: Clases desbalanceadas

Balanced Dataset



Unbalanced Dataset

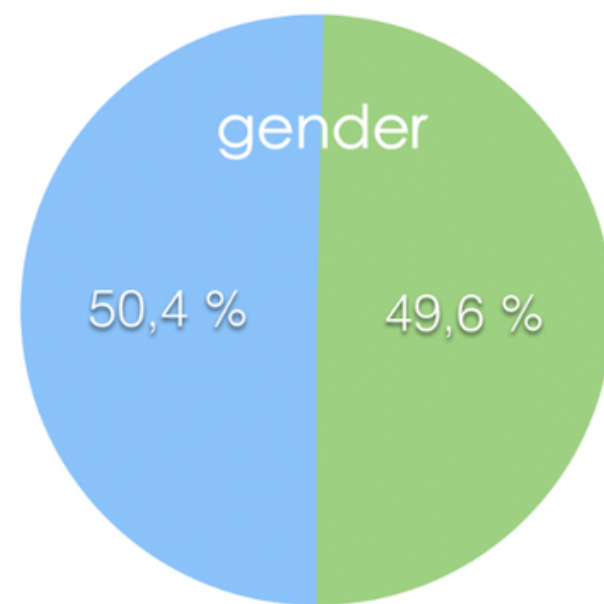


# Limitaciones de accuracy score



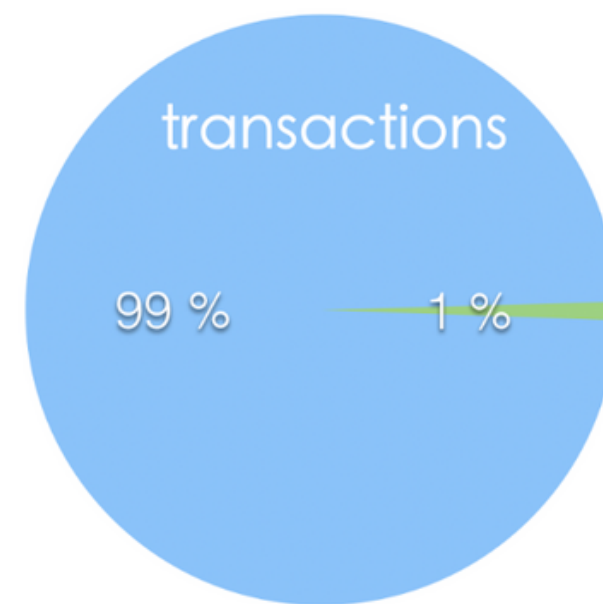
En el caso de la derecha (dataset desbalanceado) si nuestro modelo predice siempre "normal" el accuracy\_score es de 0.99, casi perfecto. Sin embargo, el modelo no sirve de nada ya que nunca identifica transacciones fraudulentas.

Balanced Dataset



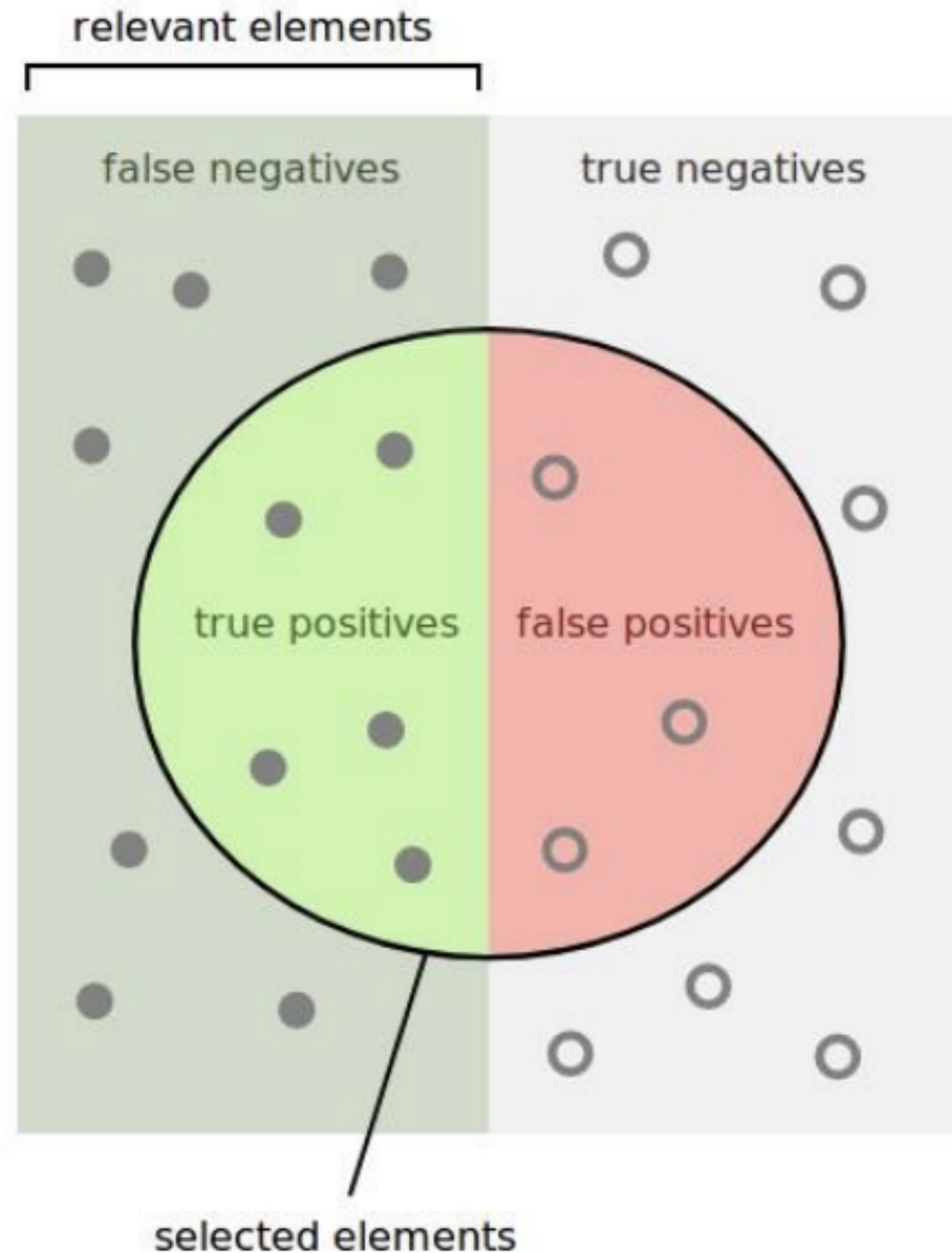
● male ● female

Unbalanced Dataset



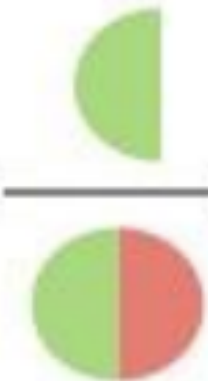
● normal ● fraudulent

# Precision - recall



How many selected items are relevant?

Precision =



$$Precision = \frac{TP}{TP+FP}$$

How many relevant items are selected?

Recall =



$$Recall = \frac{TP}{TP+FN}$$



# Precision - recall



Precision: ¿Cuántos de los items que clasificamos como verdaderos, realmente son verdaderos?

Recall: ¿Cuántos de los items que son verdaderos pudimos identificar?

Problema con estas métricas:

En general no alcanza con medir una sola ya que "compiten" entre sí.

Si decimos que todo es verdadero (todas las personas están embarazadas) nuestro recall va a ser perfecto, pero nuestra precision va a ser muy mala seguramente.

# F1 Score



Una métrica que busca un equilibrio entre precision y recall es F1 score:

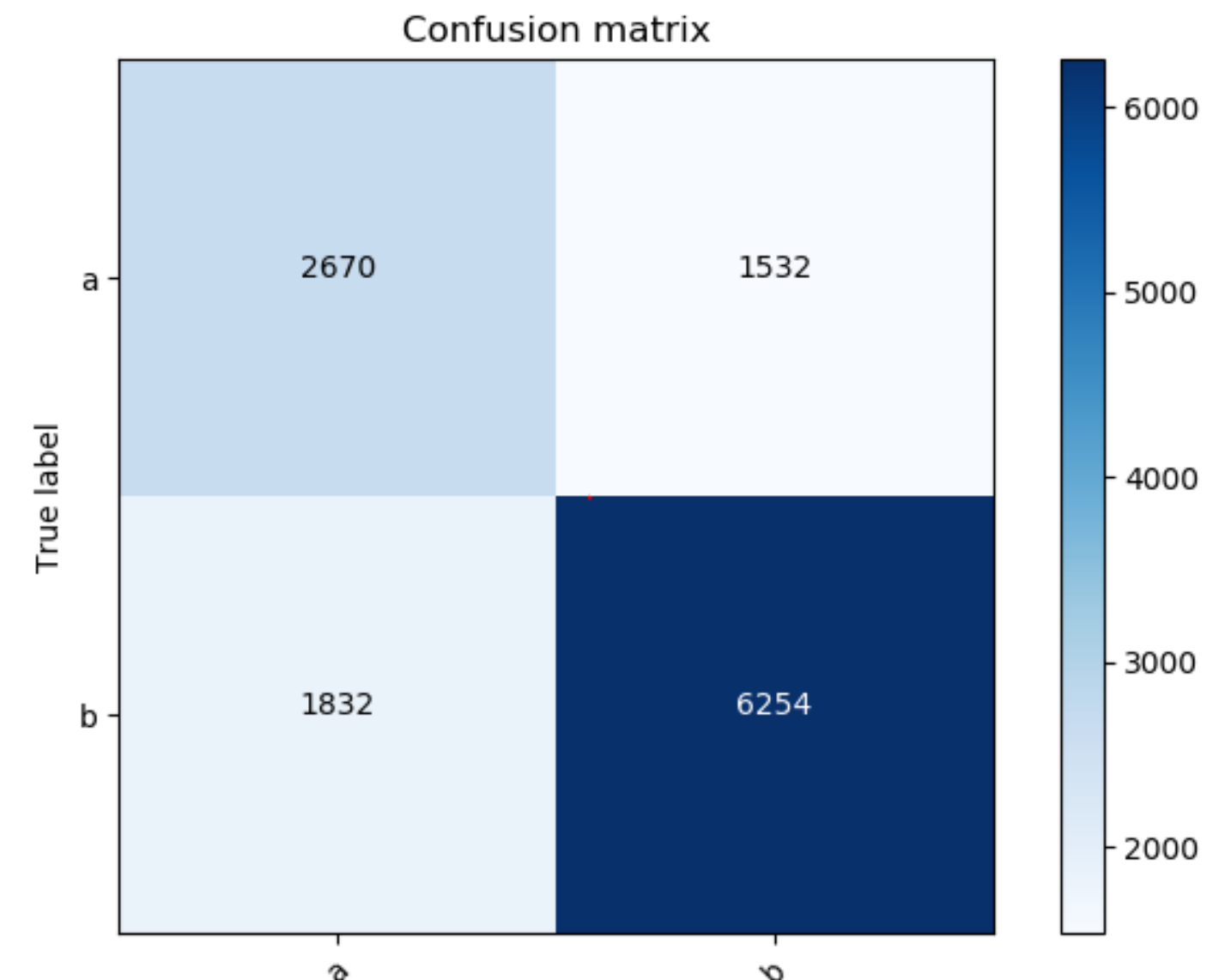
$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

# Matriz de confusión



En la matriz de confusión, vemos toda la información necesaria para calcular las métricas que definimos anteriormente:

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

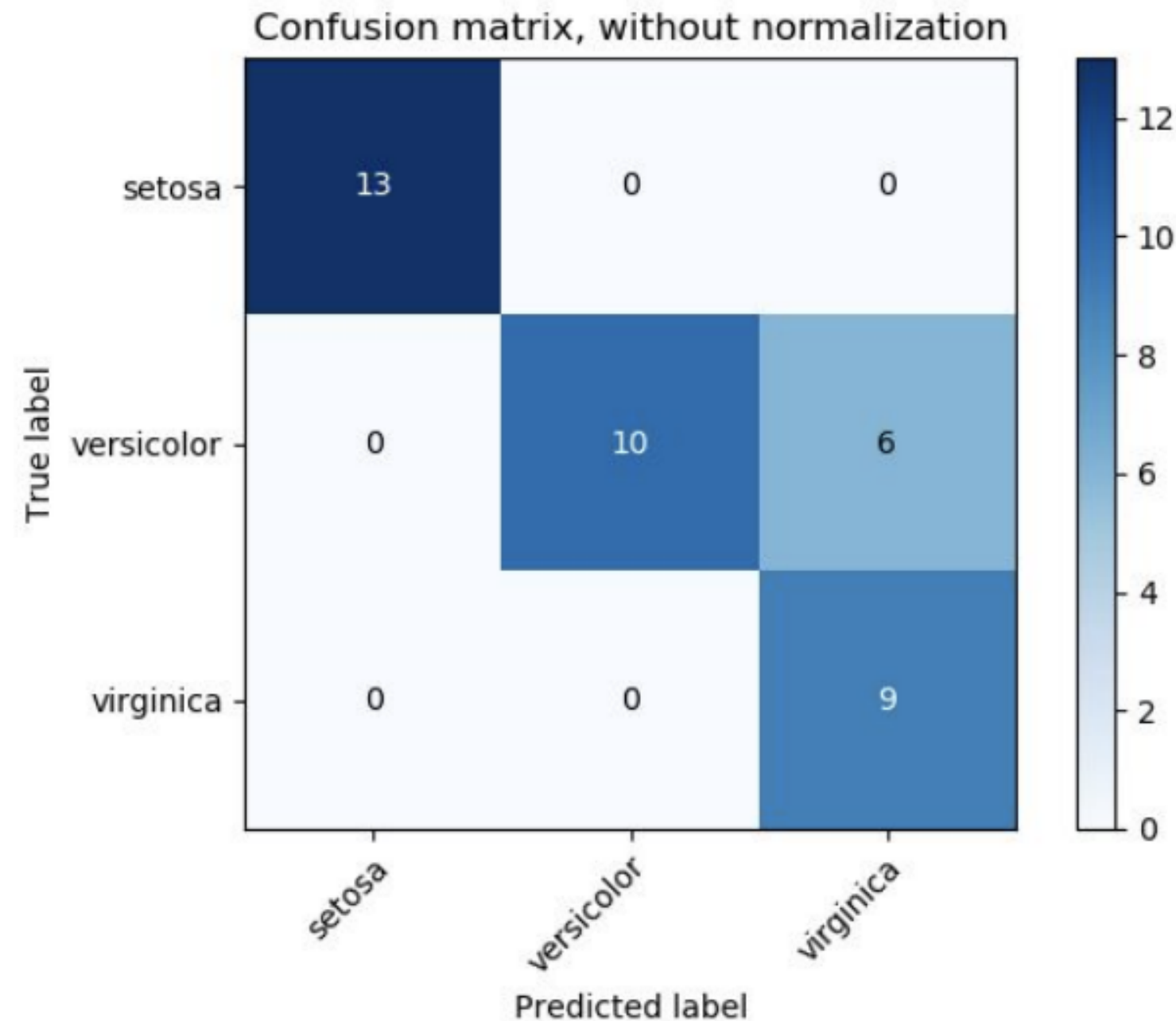


# Clasificación multiclase



¿ Cómo armamos la matriz de confusión ?

Precision & recall: por clase



**Kahoot!**

**Abrimos notebook**  
**"metricas\_clasificación.ipynb"**