



ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

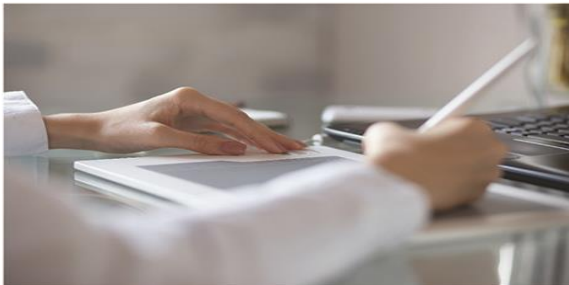
EDUCACIÓN  
PROFESIONAL

# Diplomado en Big Data y Ciencias de Datos

## Minería de Datos Matrices de Confusión

Educación Profesional - Escuela de Ingeniería UC

Sebastián Raveau



# Matriz de confusión

Dos resultados: **positivo** y **negativo**

A cuál corresponde cada posible resultado se define según el contexto de aplicación

		Referencia	
		Positivo	Negativo
Predicho	Positivo	<b>TP</b> Verdadero Positivo	<b>FP</b> Falso Positivo
	Negativo	<b>FN</b> Falso Negativo	<b>TN</b> Verdadero Negativo

# Ejemplo

Modelo KNN de riesgo crediticio, con  $K = 3$   
(visto en la sesión 5 del curso)

En este caso:

Positivo = Mala evaluación de riesgo

Negativo = Buena evaluación de riesgo

Se pueden definir al revés, simplemente cambia la interpretación de algunos indicadores

		Referencia	
		Positivo	Negativo
Predicho	Positivo	TP 22	FP 28
	Negativo	FN 58	TN 142

# Indicadores: Exactitud (*accuracy*)

Corresponde a la tasa de clasificación correcta

$$A = \frac{TP+TN}{TP+FP+FN+TN}$$

En nuestro ejemplo:

$$A = \frac{22+142}{22+28+58+142} = 0,656$$

		Referencia	
		Positivo	Negativo
Predicho	Positivo	TP 22	FP 28
	Negativo	FN 58	TN 142

# Indicadores: Tasa de No Información (*no information rate*)

Corresponde a la tasa mayoritaria  
(depende de los datos, no del modelo)

$$NIR = \frac{\max\{TP+FN; FP+TN\}}{TP+FP+FN+TN}$$

En nuestro ejemplo:

$$NIR = \frac{28+142}{22+28+58+142} = 0,680$$

		Referencia	
		Positivo	Negativo
Predicho	Positivo	TP 22	FP 28
	Negativo	FN 58	TN 142

# Indicadores: Sensibilidad (*sensitivity*)

Corresponde a la tasa de clasificación correcta de los datos positivos

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

En nuestro ejemplo:

$$Se = \frac{22}{22+58} = 0,275$$

		Referencia	
		Positivo	Negativo
Predicho	Positivo	TP 22	FP 28
	Negativo	FN 58	TN 142

# Indicadores: Especificidad (*specificity*)

Corresponde a la tasa de clasificación correcta de los datos negativos

$$Sp = \frac{TN}{TN+FP}$$

En nuestro ejemplo:

$$Sp = \frac{142}{142+28} = 0,835$$

		Referencia	
		Positivo	Negativo
Predicho	Positivo	TP 22	FP 28
	Negativo	FN 58	TN 142

# Indicadores: Exactitud Balanceada (*balanced accuracy*)

Corresponde al promedio entre sensibilidad y especificidad

$$BA = \frac{Se + Sp}{2}$$

En nuestro ejemplo:

$$BA = \frac{0,275 + 0,835}{2} = 0,555$$

		Referencia	
		Positivo	Negativo
Predicho	Positivo	TP 22	FP 28
	Negativo	FN 58	TN 142



# Indicadores: Prevalencia (*prevalence*)

Corresponde a la tasa de datos positivos  
(depende de los datos, no del modelo)

$$P = \frac{TP+FN}{TP+FP+FN+TN}$$

En nuestro ejemplo:

$$P = \frac{22+58}{22+28+58+142} = 0,320$$

		Referencia	
		Positivo	Negativo
Predicho	Positivo	TP 22	FP 28
	Negativo	FN 58	TN 142

# Indicadores: Valor de Predicción Positiva (*pos pred value*)

Corresponde a la tasa de predicciones positivas clasificadas correctamente

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

En nuestro ejemplo:

$$PPV = \frac{22+28}{22+28} = 0,440$$

		Referencia	
		Positivo	Negativo
Predicho	Positivo	TP 22	FP 28
	Negativo	FN 58	TN 142

# Indicadores: Valor de Predicción Negativa (*neg pred value*)

Corresponde a la tasa de predicciones negativas clasificadas correctamente

$$NPV = \frac{TN}{TN + FN}$$

En nuestro ejemplo:

$$NPV = \frac{142}{142+58} = 0,710$$

		Referencia	
		Positivo	Negativo
Predicho	Positivo	TP 22	FP 28
	Negativo	FN 58	TN 142

# Indicadores: Tasa de Detección (*detection rate*)

Corresponde a la tasa de verdaderos positivos, sobre el total de datos

$$DR = \frac{TP}{TP+FP+FN+TN}$$

En nuestro ejemplo:

$$DR = \frac{22}{22+28+58+142} = 0,088$$

		Referencia	
		Positivo	Negativo
Predicho	Positivo	TP 22	FP 28
	Negativo	FN 58	TN 142

# Indicadores: Prevalencia de Detección (*detection prevalence*)

Corresponde a la tasa de positivos predichos,  
sobre el total de datos

$$DP = \frac{TP+FP}{TP+FP+FN+TN}$$

En nuestro ejemplo:

$$DP = \frac{22+28}{22+28+58+142} = 0,200$$

		Referencia	
		Positivo	Negativo
Predicho	Positivo	TP 22	FP 28
	Negativo	FN 58	TN 142

# Estadígrafos: Exactitud al azar (*random accuracy*)

Corresponde a la exactitud de un modelo que clasifica al azar

$$p_T = \frac{TP+FN}{TP+FP+FN+TN} ; p_N = \frac{TP+FP}{TP+FP+FN+TN}$$

$$RA = p_T \cdot p_N + (1 - p_T) \cdot (1 - p_N)$$

En nuestro ejemplo:

$$RA = 0,320 \cdot 0,200 + 0,680 \cdot 0,800 = 0,608$$

		Referencia	
		Positivo	Negativo
Predicho	Positivo	TP 22	FP 28
	Negativo	FN 58	TN 142

# Estadígrafos: Kappa

Compara la exactitud del modelo con la exactitud al azar

$$\kappa = \frac{A - RA}{1 - RA}$$

En nuestro ejemplo:

$$\kappa = \frac{0,656 - 0,608}{1 - 0,608} = 0,122$$

		Referencia	
		Positivo	Negativo
Predicho	Positivo	TP 22	FP 28
	Negativo	FN 58	TN 142