Maestría en Inteligencia Artificial Aplicada

Matriz de Confusión

Inteligencia Artificial y Aprendizaje Automático

Dr. Luis Eduardo Falcón Morales

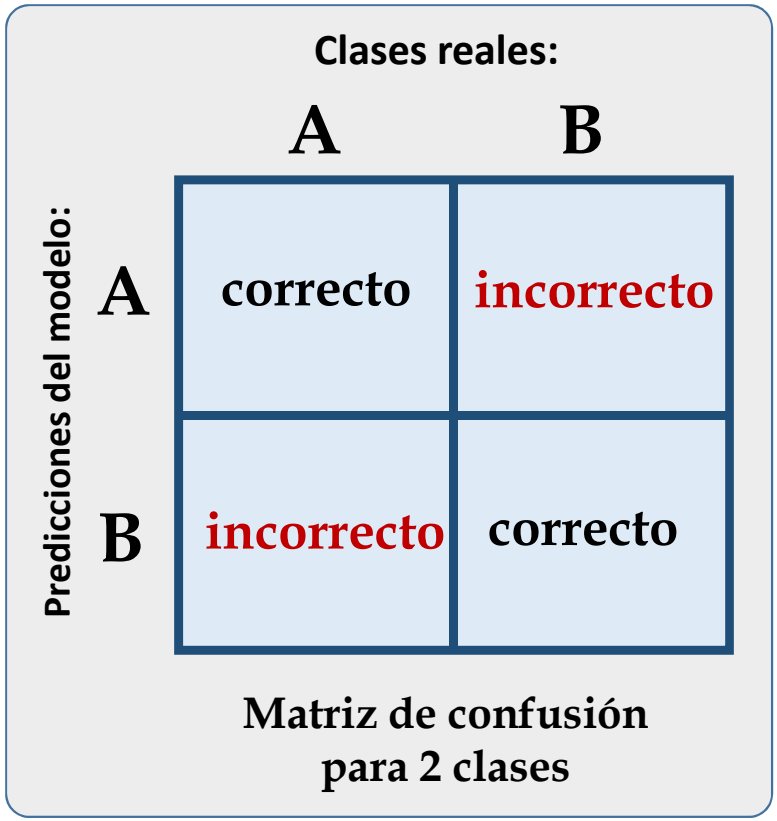
Matriz de Confusión o Tablas de Contingencia

- La matriz de confusión es una de las mejores herramientas de aprendizaje supervisado que visualiza mediante una matriz, qué tan bien se llevan a cabo las predicciones en un problema de clasificación.
- La matriz de confusión compara los datos reales contra los datos pronosticados por el modelo.
- En el área de Estadística se les llama tablas de contingencia y han sido utilizadas desde principios del siglo XX.
- En el área de aprendizaje supervisado se le llama usualmente matriz de confusión, en el sentido de qué tanto se confunde el modelo predictivo utilizado al hacer sus inferencias sobre las distintas clases.
- En ocasiones también se le llama simplemente matriz de los errores.
- Algunos autores manejan de manera indistinta el nombre de matriz de confusión y la tabla de contingencia.

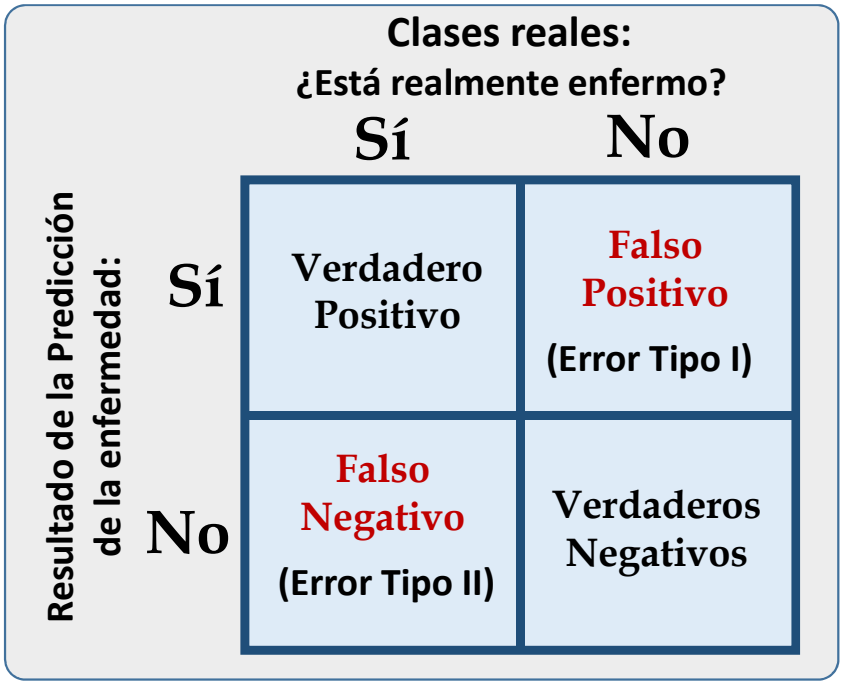
Problema Binario: Falsos Positivos y Falsos Negativos

A la clase de interés la llamaremos la clase positiva.

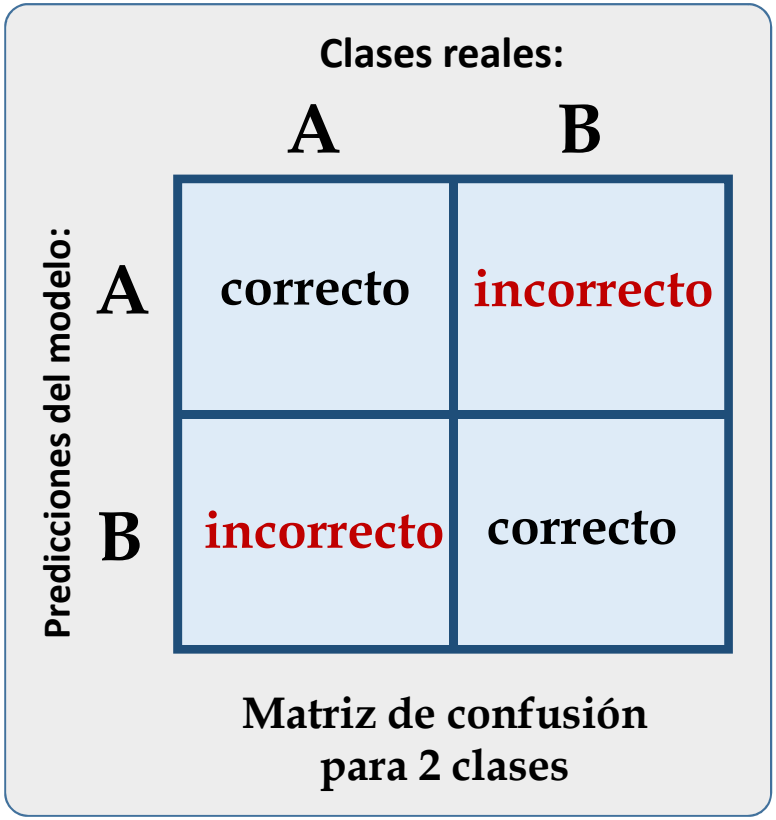
A la otra clase o clases las llamaremos clase o clases negativas.



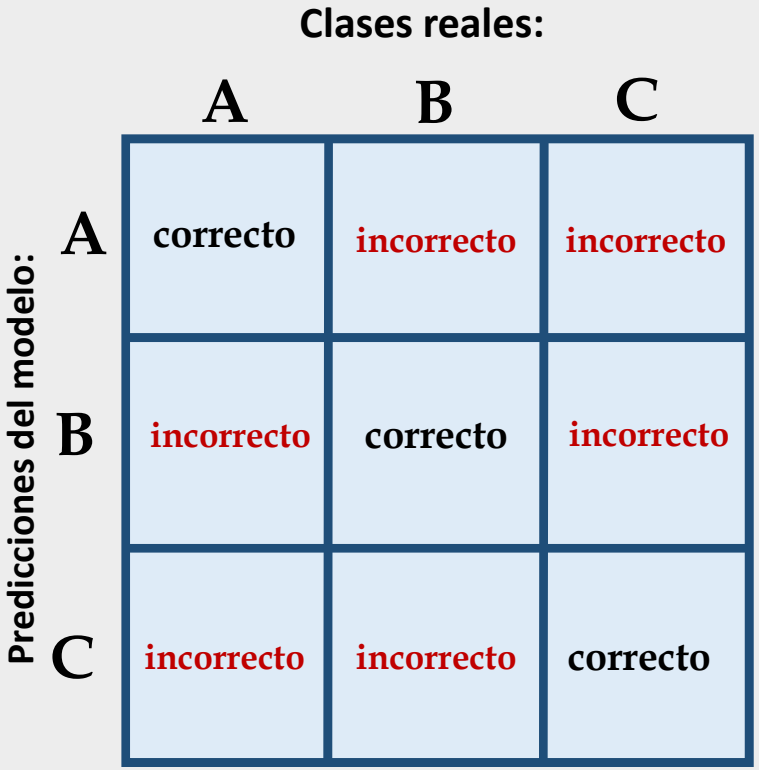
Matriz de confusión para 2 clases



¿qué error cuesta más? ¿el Tipo I o el II?

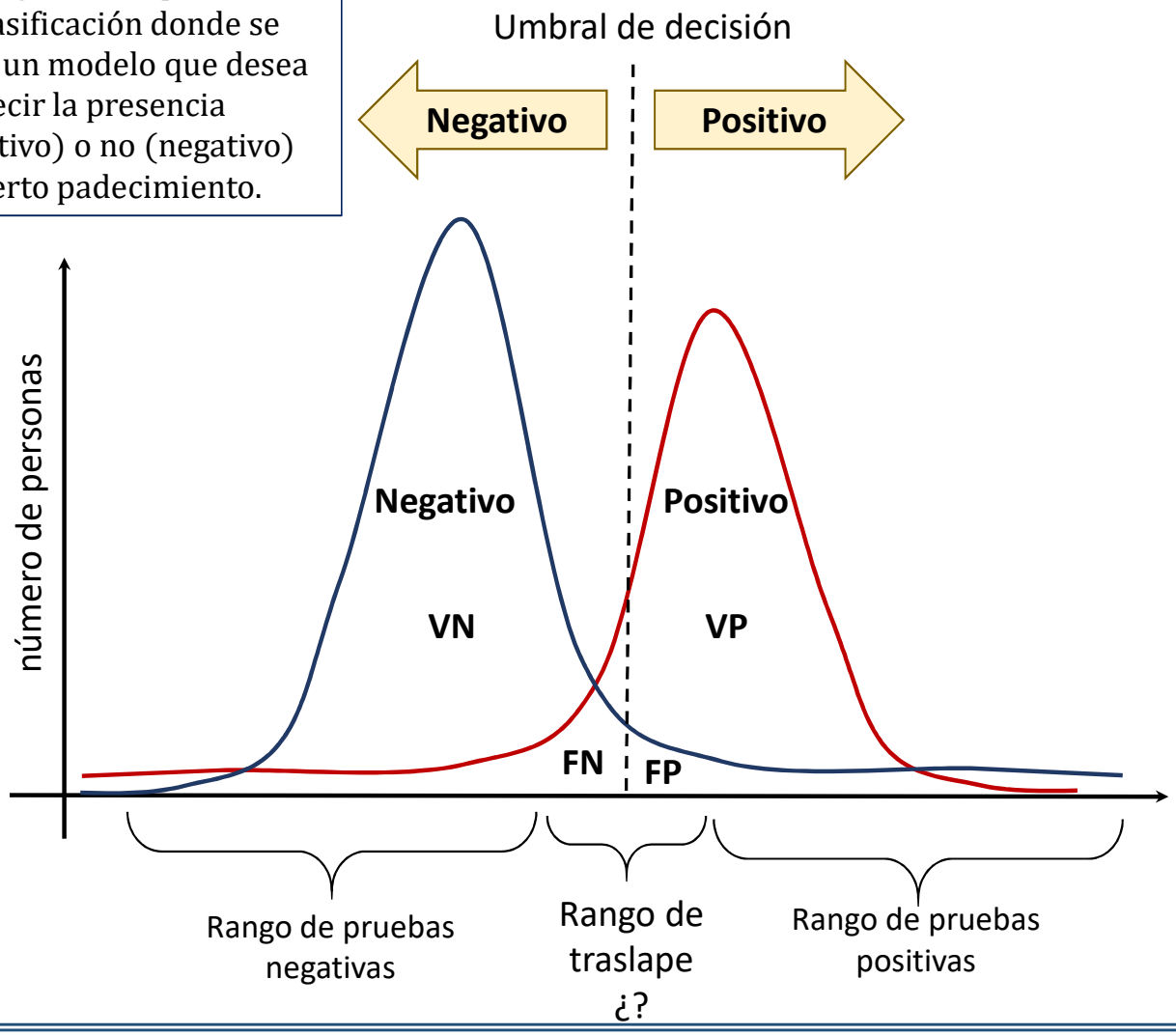


Matriz de confusión
para 2 clases



Matriz de confusión
para 3 clases

Supongamos un problema de clasificación donde se tiene un modelo que desea predecir la presencia (positivo) o no (negativo) de cierto padecimiento.

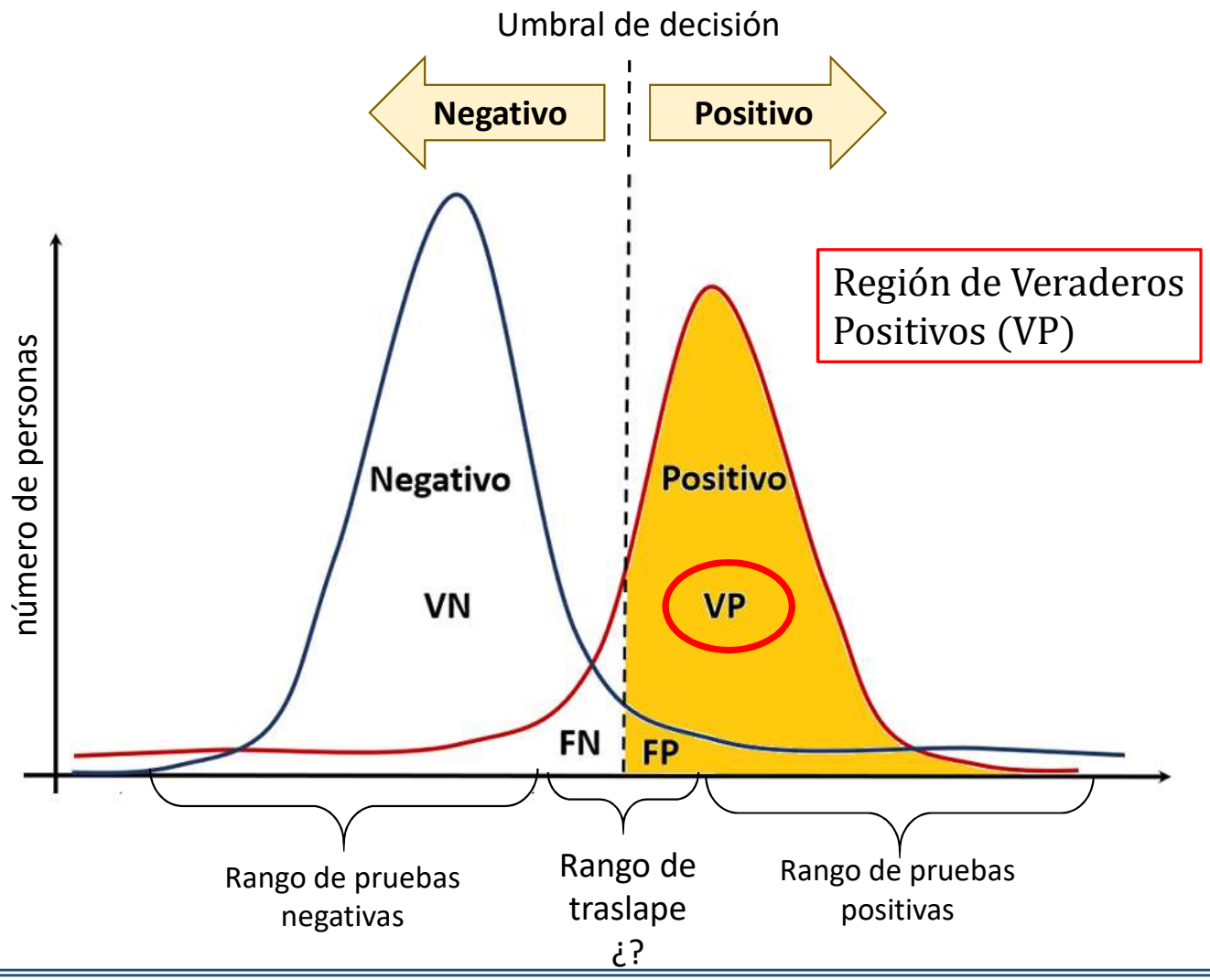


	¿Positivo?	
	Clase real: Sí	Clase real: No
Predicción: Sí	VP	FP
Predicción: No	FN	VN

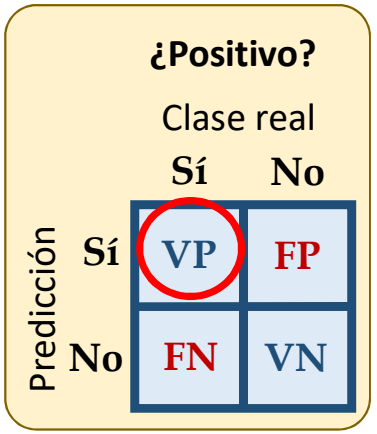
Relación con los conceptos en Estadística:

FP : Error Tipo I (alfa)
FN : Error Tipo II (beta)
Potencia : 1 - beta

En analogía al concepto de prueba de hipótesis de Estadística, podemos visualizar los términos de la matriz de confusión en las regiones de las curvas de distribución de los datos de entrenamiento mostradas.

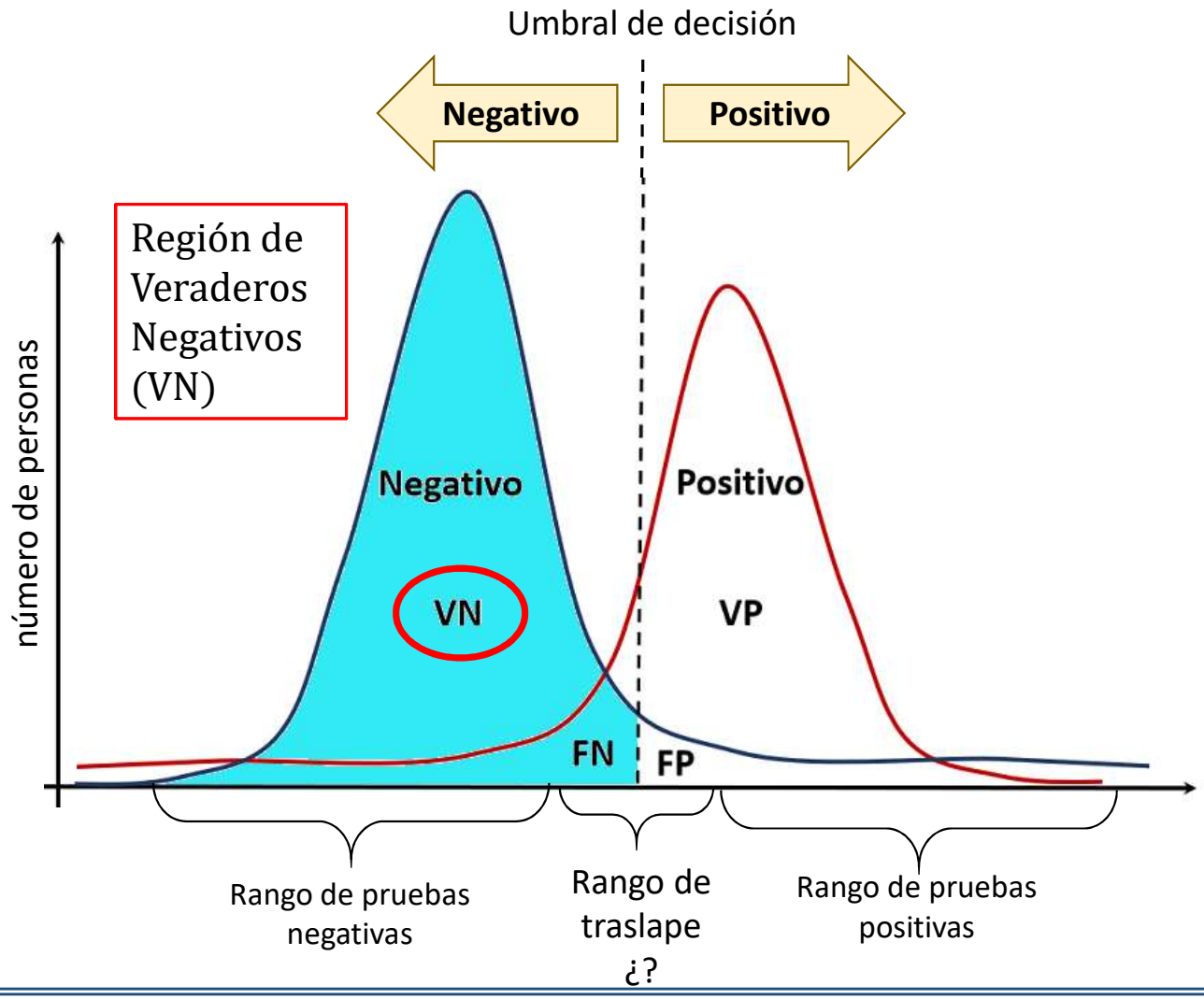


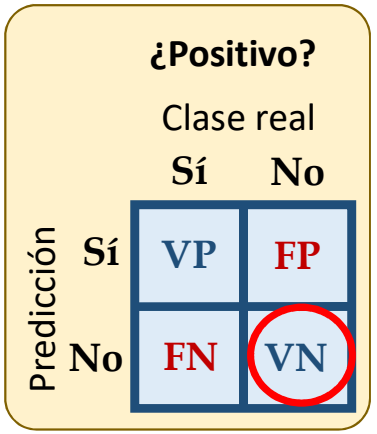
Región de Veraderos Positivos (VP)



Relación con los conceptos en Estadística:

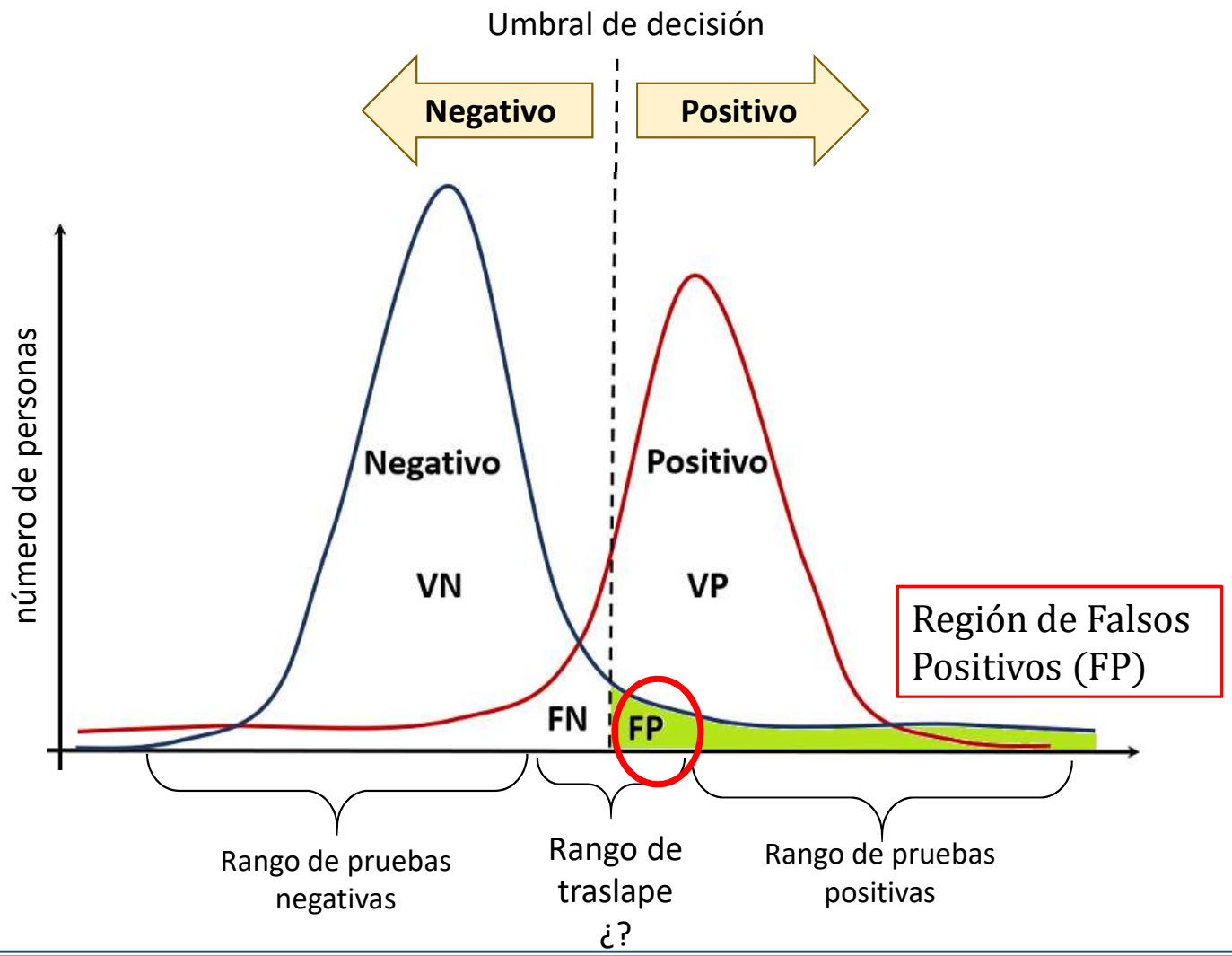
FP : Error Tipo I (alfa)
FN : Error Tipo II (beta)
Potencia : 1 - beta

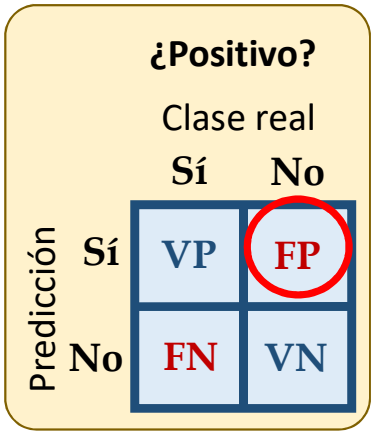




Relación con los conceptos en Estadística:

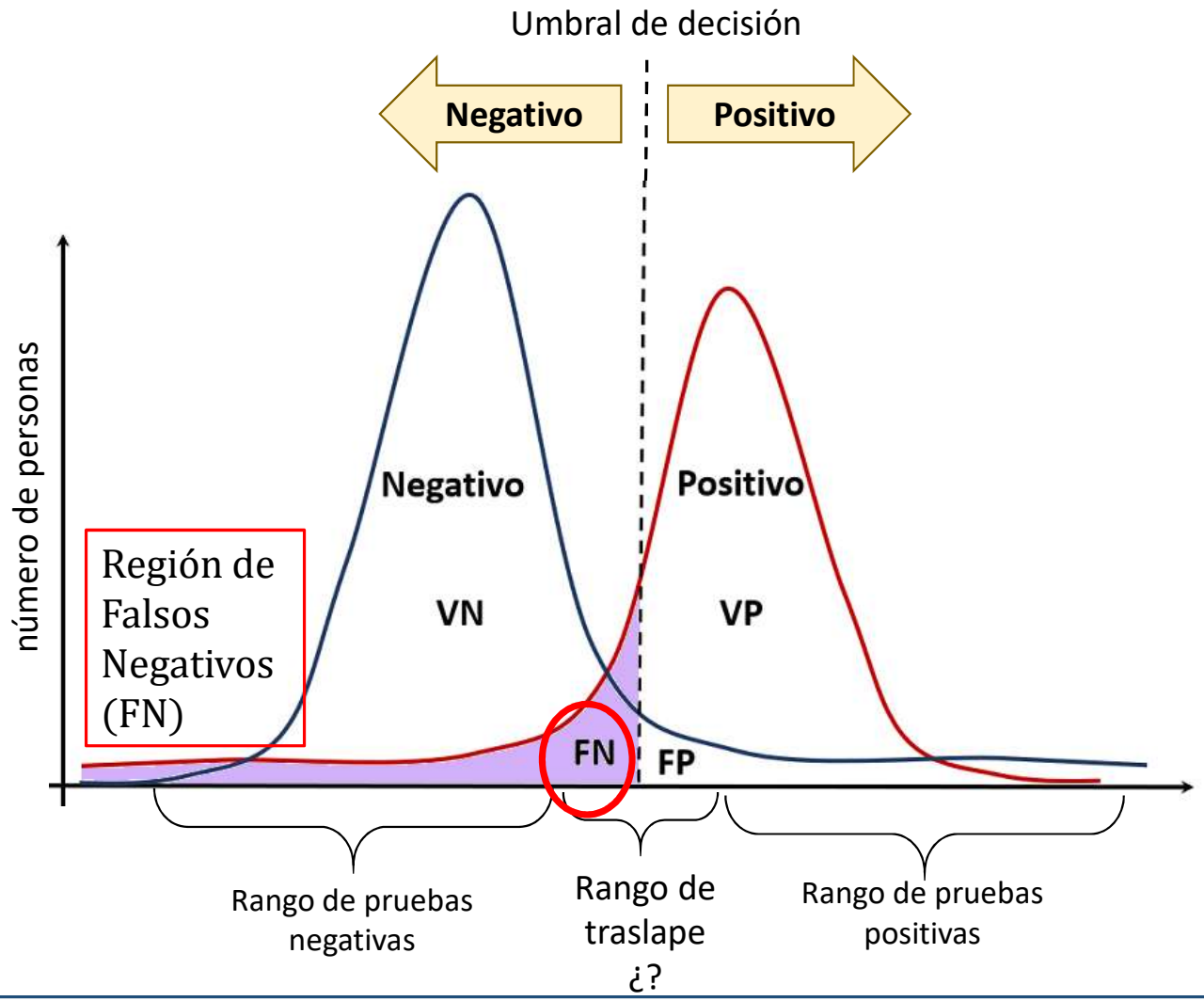
FP : Error Tipo I (alfa)
FN : Error Tipo II (beta)
Potencia : 1 - beta

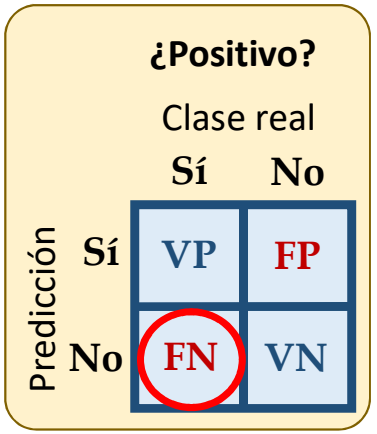




Relación con los conceptos en Estadística:

FP : Error Tipo I (alfa)
FN : Error Tipo II (beta)
Potencia : 1 - beta





Relación con los conceptos en Estadística:

FP : Error Tipo I (alfa)
FN : Error Tipo II (beta)
Potencia : 1 - beta

Matriz de Confusión

A partir de los valores de la matriz de confusión podemos generar otra serie de valores y métricas que nos permitirán medir el desempeño de un modelos de distintas maneras.

		Clases reales		**Totales:**
		A clase positiva	**B** clase negativa	
Predicciones	**A** clase positiva	VP $\frac{VP}{n_{r1}}$ $\frac{VP}{n_{c1}}$ $\frac{VP}{n_T}$	FP $\frac{FP}{n_{r1}}$ $\frac{FP}{n_{c2}}$ $\frac{FP}{n_T}$	$n_{r1} = VP + FP$ $\frac{n_{r1}}{n_T}$
	B clase negativa	FN $\frac{FN}{n_{r2}}$ $\frac{FN}{n_{c1}}$ $\frac{FN}{n_T}$	VN $\frac{VN}{n_{r2}}$ $\frac{VN}{n_{c2}}$ $\frac{VN}{n_T}$	$n_{r2} = FN + VN$ $\frac{n_{r2}}{n_T}$
Totales:		$n_{c1} = VP + FN$ $\frac{n_{c1}}{n_T}$	$n_{c2} = FP + VN$ $\frac{n_{c2}}{n_T}$	$n_T = n_{r1} + n_{r2}$ $= n_{c1} + n_{c2}$

Hasta ahora, la manera más sencilla de medir el desempeño de un algoritmo de aprendizaje supervisado en un problema de clasificación, es contabilizando el total de aciertos entre el total de predicciones:

$$\%\ de\ éxito \quad = \frac{número\ de\ aciertos}{total\ de\ predicciones} \times 100\%$$

Esta medida nos da un porcentaje de casos en los cuales el algoritmo hace predicciones correctas.

A este valor se le suele llamar exactitud (accuracy), pero existen otras muchas métricas que veremos a continuación. Además, veremos que en general conviene expresarlas en términos de los valores VP, VN, FP y FN de la matriz de confusión.

Estadísticos

Existen una gran variedad de métricas para medir el desempeño de un clasificador.

Las siguientes son algunas de las principales:

- Exactitud
- Sensibilidad
- Especificidad
- Precisión
- Exhaustividad
- valor-F
- curva ROC
- Área bajo la curva

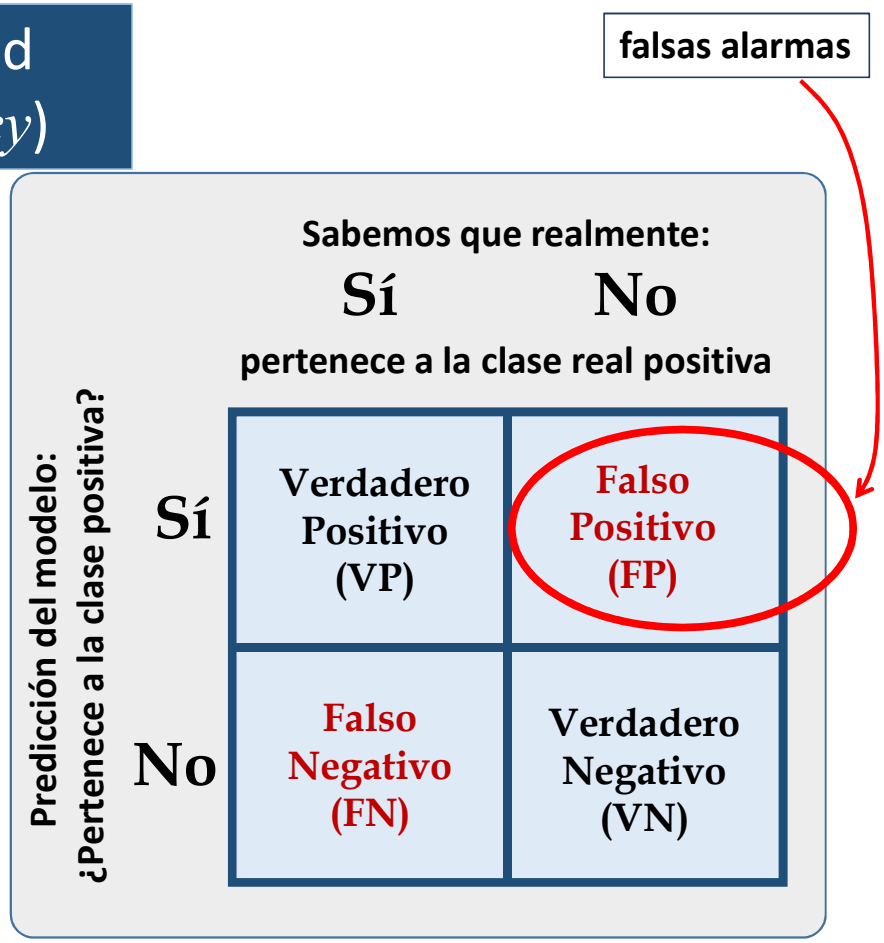
Exactitud (*accuracy*)

Exactitud o tasa de éxito:

$$\underset{(accuracy)}{exactitud} = \frac{VP + VN}{VP + VN + FP + FN}$$

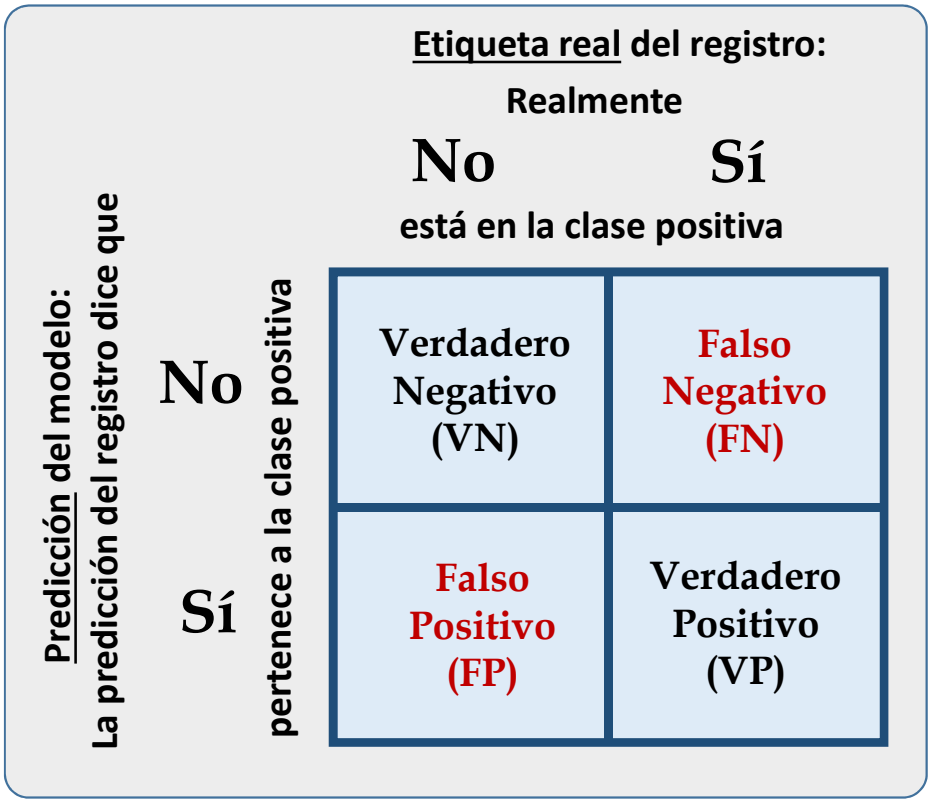
Tasa de error, es decir, tasa de clases mal clasificadas:

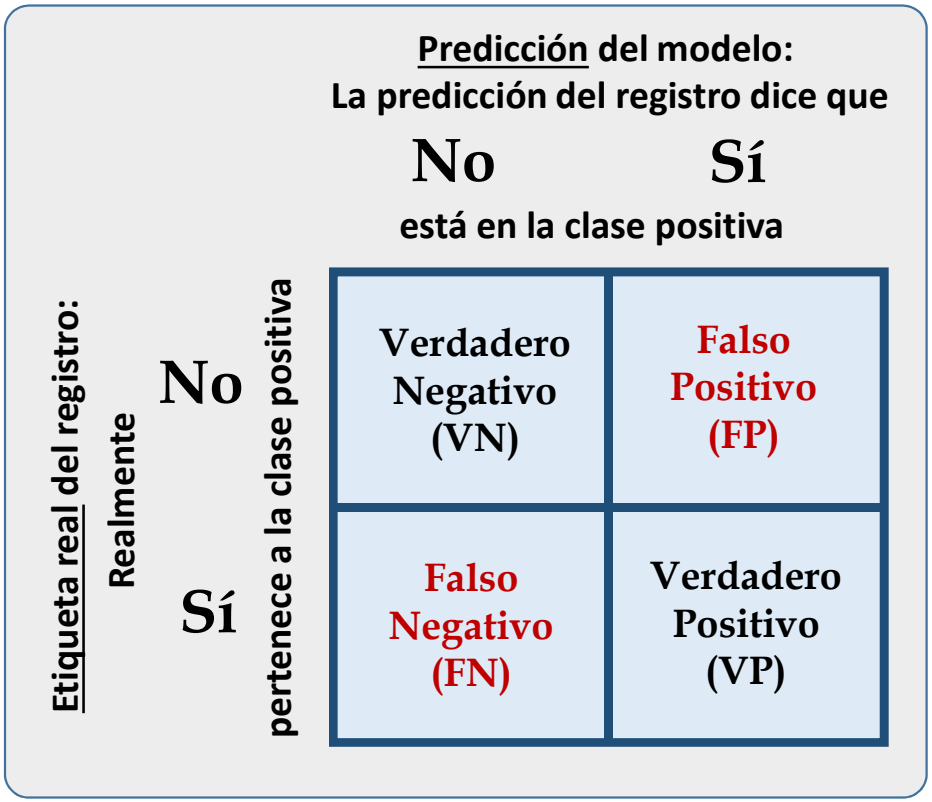
$$\underset{error}{tasa\ de} = 1 - exactitud$$



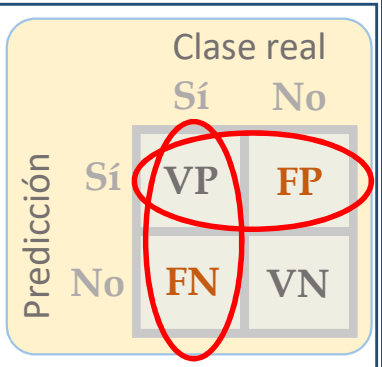
Matriz de Confusión

El orden en que se muestren los valores reales y los de predicción en las columnas y los renglones en la matriz de confusión, es completamente arbitrario. Por ello deberás siempre observar con detalle la manera en que autores y librerías los despliegan. Así, las siguientes dos matrices son completamente equivalentes.





Precisión y Exhaustividad



Son métricas sobre qué tan relevante son los resultados dados por un modelo, con respecto a los verdaderos positivos.

Precisión (precision)

$$precision = \frac{VP}{VP + FP}$$

tasa de Predicciones positivas.

Cuando un modelo predice que un dato pertenece a la clase positiva, ¿qué porcentaje de acierto hay en ello?

Y definimos la **tasa de falsas alarmas**:

$$1 - precisión = \frac{FP}{VP+FP}$$

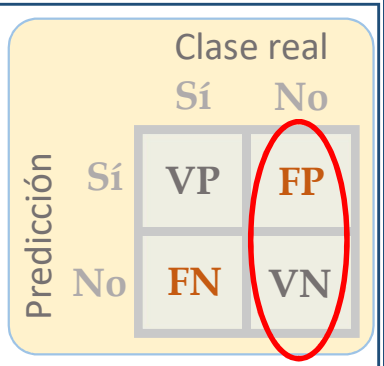
Exhaustividad / Sensibilidad (recall / sensitivity)

$$\begin{matrix} sensibilidad \\ exhaustividad \end{matrix} = \frac{VP}{VP + FN}$$

tasa de Verdaderos positivos.

Es decir, ¿qué porcentaje de los elementos de la clase positiva son pronosticados correctamente?

Especificidad



Esta métrica ayuda a dar seguimiento de los verdaderos negativos, estén estos bien o mal clasificados por nuestro estimador.

Especificidad
(specificity)

$$especificidad = \frac{VN}{FP + VN}$$

O **tasa de verdaderos negativos.** Es decir, ¿qué porcentaje de los elementos de la clase negativa son pronosticados correctamente?

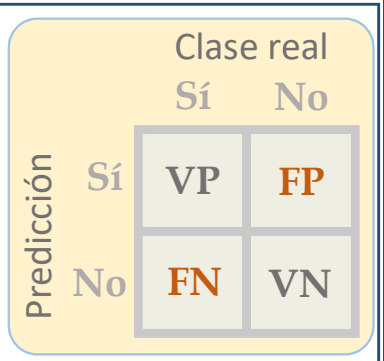
La **tasa de falsos positivos** es: $1 - especificidad = \frac{FP}{VN+F}$

NOTA: Esta métrica se usará junto con la sensibilidad (recall) para generar las curvas ROC

Ejemplo:

Supongamos que tenemos una aplicación para detectar correo spam, y sabemos que últimamente se han comportado como se muestra en la tabla:

Predicciones \ Valores Reales	Spam	Ham
Spam	412	256
Ham	17	2137



Uno debe hacer ajustes con respecto a qué tan severo quiero que sea la detección de *spams*, sin castigar demasiado a los correos no *spam* bien clasificados.

Entonces, considerando Spam como la clase positiva, tendremos:

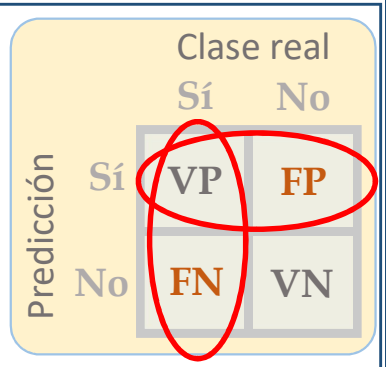
$$sensibilidad = \frac{412}{412 + 17} = 0.9604$$
$$exhaustividad$$

Es decir, 96.04% de los correos spam son detectados correctamente y un 4.96% de los correos spam se clasifican erróneamente como Ham.

$$especificidad = \frac{2137}{2137 + 256} = 0.8930$$

Así, un 89.3% de los correos no spam son correctamente clasificados, y un 10.7% de ellos se consideraron erróneamente como spam.

valor_F (f1_score)



La media armónica H de dos números a y b, se define como $\frac{1}{H}=\frac{1}{2}\left(\frac{1}{a}+\frac{1}{b}\right)$

La media armónica siempre es menor o igual a la media aritmética.

En particular a la media armónica de la precisión y la exhaustividad se le llama f1_score:

$$f1\text{ -score}=\frac{2VP}{2VP+FP+FN}$$

Valores entre 0 (menor exactitud) y 1 (mayor exactitud).

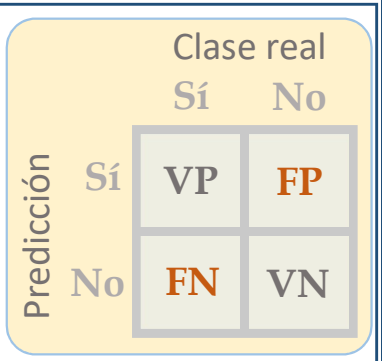
No aplica y por lo tanto no se recomienda si alguno de los valores es cero.

También se le llama valor_F, medida_F, F_value, F_measure, F_score, $F1$.

Utilicemos la exactitud (accuracy) cuando se tienen clases balanceadas y los VP y VN son lo principal a monitorear.

Utilicemos la f1-score cuando se tienen clases desblanceadas y algunos de los valores FP o FN son prioridad para monitorear.

Ejemplo: Continuemos con los datos del mismo ejemplo anterior sobre una aplicación para detectar correo spam:



Valores Reales

Predicciones	Spam	Ham
Spam	412	256
Ham	17	2137

Entonces, considerando Spam como la clase positiva, tendremos:

$$precisión = \frac{412}{412 + 256} = 0.6168$$

Es decir, el 61.68% de las veces que mi aplicación dice que detectó un correo spam, esta es una afirmación correcta.

Y por lo tanto, el 38.32% $(1 - 0.6168)$ de las veces, son falsas alarmas.

Y previamente encontramos que exhaustividad es igual a 0.9604

Entoncds la media armónica entre precisión y exhaustividad nos da:

$$f1\text{-}score = 0.7511$$