

# MATEMÁTICAS Y DATOS X... in a nutshell.

## Coeficiente de Otsuka-Ochiai



Comparar dos registros en una base de datos es indispensable para la minería de datos y algunos algoritmos de ML

Supongamos que A y B son dos registros que contienen o no información sobre **D** temas.

$$A = \{a_1, a_2, \dots, a_D\}$$

$$B = \{b_1, b_2, \dots, b_D\}$$

Calculemos el número **d** de temas en común entre los registros A y B.

$$A \cap B = \{c_1, c_2, \dots, c_d\}$$

Después dividamos sobre **a** el número de temas mencionados en A multiplicado por **b** el número de temas mencionados en B.

$$\frac{d}{a \cdot b}$$

$$\frac{d}{\sqrt{a \cdot b}}$$

Para finalizar calculamos la raíz cuadrada para no contar repetidos.




El coeficiente de Otsuka-Ochiai es una versión conjuntista del cálculo de la medida del coseno comúnmente utilizada en Ciencia de Datos para calcular la similitud entre dos registros.

De hecho para vectores representados exclusivamente por variables binarias, ambas cantidades coinciden.

✉ [info@colegio-bourbaki.com](mailto:info@colegio-bourbaki.com)

☎ +52 999 267 7399

 Colegio de Matemáticas Bourbaki

