

# **DATA SCIENCE** **APLICADO A** **NEGOCIOS**



- **El procesamiento de los lenguajes naturales se puede usar para crear modelos predictivos que lleven a cabo análisis del sentimiento basándose en posts y valoraciones de redes sociales y que predigan si el cliente está contento o no.**
- **Los procesadores del lenguaje natural trabajan convirtiendo palabras a números y entrenando modelos de Machine Learning para hacer predicciones.**
- **De este modo, podemos saber si nuestros clientes están contentos o no sin tener que revisar manualmente un número inmenso de tweets o de valoraciones!**



- **Trabajas como data scientist en una corporación multinacional.**
- **El departamento de relaciones públicas ha recolectado una gran cantidad de datos sobre sus clientes como valoraciones de sus productos en internet.**
- **Basándose en esas valoraciones (en formato de texto), el equipo busca predecir si sus clientes están satisfechos o no con su producto.**



TOKENIZACIÓN

This is the first document.  
This document is the second document.  
And this is the third one.  
Is this the first document?



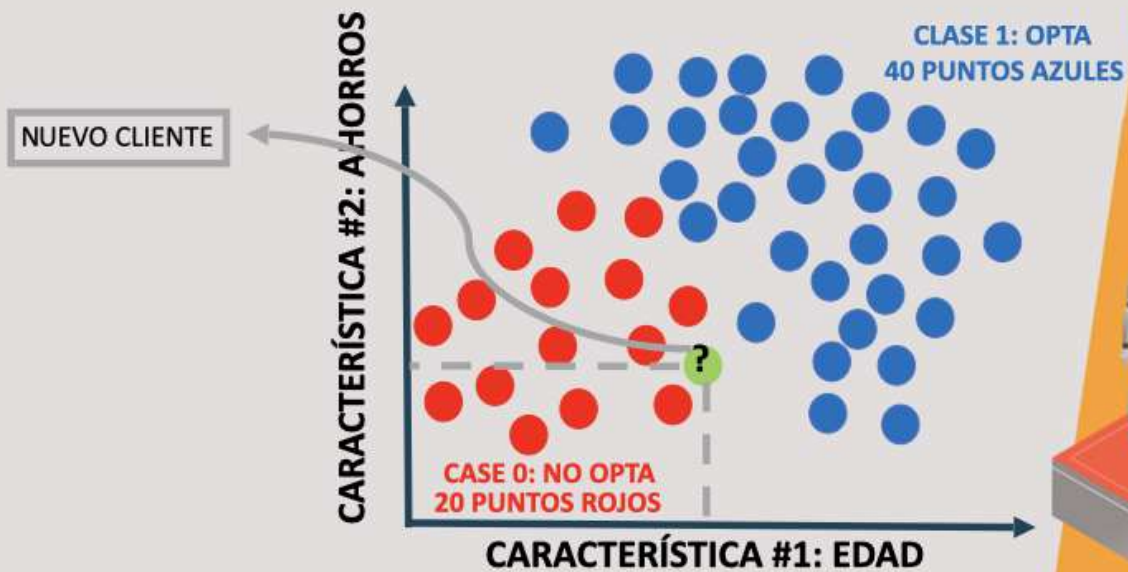
[[0 1 1 1 0 0 1 0 1]  
[0 2 0 1 0 1 1 0 1]  
[1 0 0 1 1 0 1 1 1]  
[0 1 1 1 0 0 1 0 1]]

	'and'	'document'	'first'	'is'	'one'	'second'	'the'	'third'	'this'
Training Sample #1	0	1	1	1	0	0	1	0	1
Training Sample #2	0	2	0	1	0	1	1	0	1
Training Sample #3	1	0	0	1	1	0	1	1	1
Training Sample #4	0	1	1	1	0	0	1	0	1



## NAÏVE BAYES: INTUICIÓN

- Naïve Bayes es una técnica de clasificación basada en el Teorema de Bayes.
- Supongamos que eres un data scientist que trabaja en un banco de NYC y quieres clasificar si un nuevo cliente puede obtener un préstamo del banco o no.
- Las características del cliente son su edad y ahorros.



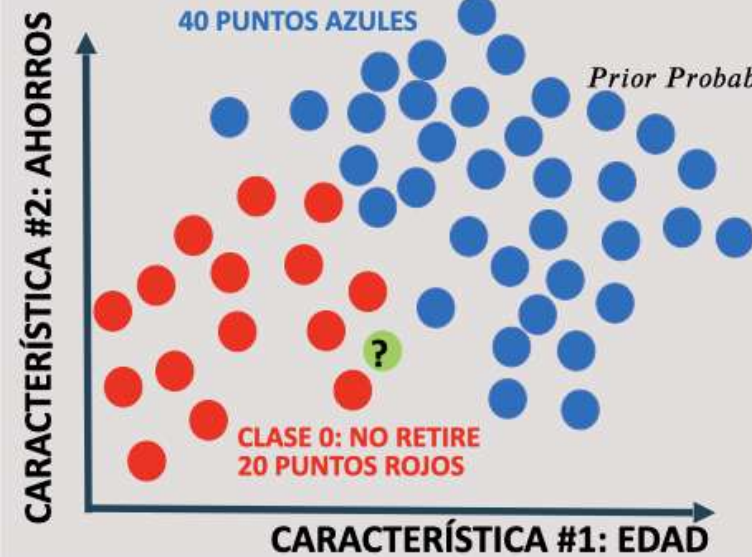


## NAÏVE BAYES: 1. PROBABILIDAD A PRIORI

- Los puntos se pueden clasificar como ROJO o AZUL y nuestro trabajo es clasificar los nuevos puntos como ROJO o AZUL.
- Probabilidad a Priori: Como tenemos más puntos AZULES que ROJOS, podemos asumir que hay el doble de probabilidades de que el nuevo punto sea AZUL que no ROJO.

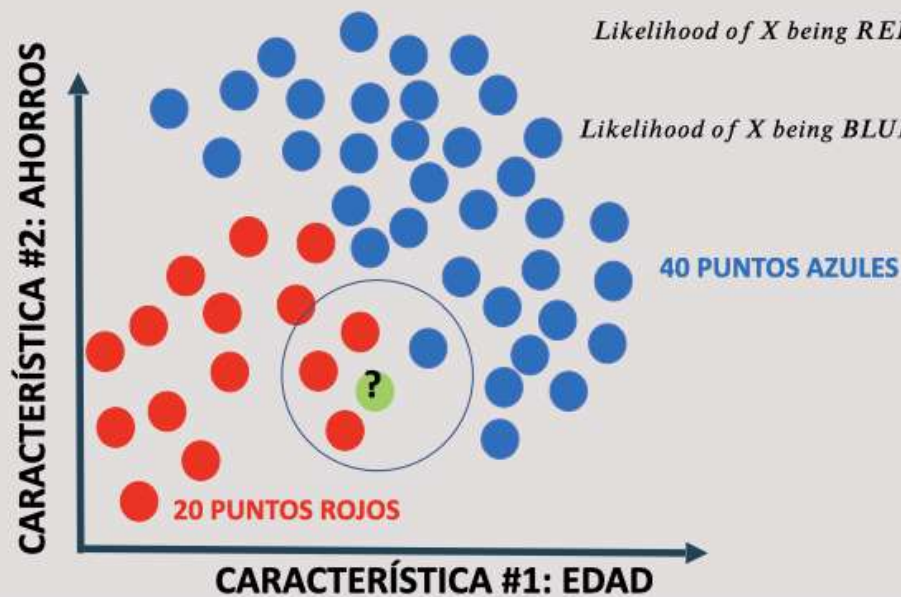
$$\text{Prior Probability for RED} = \frac{\text{Number of RED Points}}{\text{Total Number of Points}} = \frac{20}{60}$$

$$\text{Prior Probability for BLUE} = \frac{\text{Number of BLUE Points}}{\text{Total Number of Points}} = \frac{40}{60}$$



## NAÏVE BAYES: 2. VEROSIMILITUD

- Para el nuevo punto, si hay más PUNTOS AZULES en un entorno del mismo, es más probable que se clasifique como AZUL.
- Dibujamos pues un círculo alrededor del punto
- Luego calculamos cuantos puntos del círculo corresponden a cada CLASE



$$\text{Likelihood of } X \text{ being RED} = \frac{\text{Number of RED Points in vicinity}}{\text{Total Number of RED Points}} = \frac{3}{20}$$

$$\text{Likelihood of } X \text{ being BLUE} = \frac{\text{Number of BLUE Points in vicinity}}{\text{Total Number of BLUE Points}} = \frac{1}{40}$$



### NAÏVE BAYES: 3. PROBABILIDAD A POSTERIORI

- Combinamos la probabilidad a priori y la verosimilitud para crear la probabilidad a posteriori
- Probabilidad a priori: sugiere que X debería ser clasificado como BLUE porque hay el doble de PUNTOS AZULES.
- Verosimilitud: sugiere que X debería ser ROJO porque hay más PUNTOS ROJOS en un entorno de X.
- La regla de Bayes combina ambos para crear una probabilidad a posteriori.

$$\text{Posterior Probability of } X \text{ being RED} = \text{Prior Probability of RED} * \text{Likelihood of } X \text{ being RED} = \frac{20}{60} * \frac{3}{20} = \frac{1}{20}$$
$$\text{Posterior Probability of } X \text{ being BLUE} = \text{Prior Probability of BLUE} * \text{Likelihood of } X \text{ being BLUE} = \frac{40}{60} * \frac{1}{40} = \frac{1}{60}$$

CARACTERÍSTICA #2: AHORROS



CARACTERÍSTICA #1: EDAD

X SE CLASIFICA COMO ROJO (NO OPTA) YA QUE TIENE MAYOR PROBABILIDAD A POSTERIORI





## NAÏVE BAYES: UN POCO DE MATEMÁTICAS

VEROSIMILITUD

PROBABILIDAD A  
PRIORI DE OPTAR

$$P(Retire | X) = \frac{P(X | Retire) * P(Retire)}{P(X)}$$

PROBABILIDAD MARGINAL

- Naïve Bayes es una técnica de clasificación basada en el Teorema de Bayes
- $X$ : La característica de un nuevo cliente: edad y ahorros
- $P(Retire | X)$ : probabilidad de que el cliente opte al préstamo en función de sus características: edad y ahorro.
- $P(Retire)$ : Probabilidad a priori de optar al préstamo, sin conocimiento previo.
- $P(X | Retire)$ : verosimilitud
- $P(X)$ : verosimilitud marginal, la probabilidad de que cualquier punto caiga en el círculo.



## NAÏVE BAYES: UN POCO DE MATEMÁTICAS

VEROSIMILITUD

PROBABILIDAD A  
PRIORI DE OPTAR

$$P(Retire | X) = \frac{P(X | Retire) * P(Retire)}{P(X)}$$

PROBABILIDAD MARGINAL

$$P(Retire) = \frac{\# \text{ of Retiring}}{\text{Total points}} = 40/60$$

$$P(X | Retire) = \frac{\# \text{ of similar observations for retiring}}{\text{Total \# retiring}} = 1/40$$

$$P(X) = \frac{\# \text{ of Similar observations}}{\text{Total \# Points}} = 4/60$$

$$P(Retire | X) = \frac{\frac{40}{60} * \frac{1}{40}}{\frac{4}{60}} = \frac{1/60}{4/60} = 0.25$$



**NAÏVE BAYES: EVALUACIÓN, CALCULA  
LA PROBABILIDAD DE NO OPTAR AL  
PRÉSTAMO (CLASE ROJA)**

$$P(\textit{No Retire} \mid X) = ?$$



## NAÏVE BAYES: EVALUACIÓN, CALCULA LA PROBABILIDAD DE NO OPTAR AL PRÉSTAMO (CLASE ROJA)

VEROSIMILITUD

$$P(\text{No Retire} | X) = \frac{P(X | \text{No Retire}) * P(\text{No Retire})}{P(X)}$$

PROBABILIDAD A PRIORI DE NO OPTAR

CLASE 1: OPTA  
40 PUNTOS AZULES

PROBABILIDAD MARGINAL

$$P(\text{No Retire}) = \frac{\# \text{ of No Retiring}}{\text{Total points}} = 20/60$$

$$P(X | \text{No Retire}) = \frac{\# \text{ of similar observations for No retiring}}{\text{Total \# no retiring}} = 3/20$$

$$P(X) = \frac{\# \text{ of Similar observations}}{\text{Total \# Points}} = 4/60$$

$$P(\text{No Retire} | X) = \frac{\frac{20}{60} * \frac{3}{20}}{\frac{4}{60}} = \frac{3/60}{4/60} = 0.75$$

CARACTERÍSTICA #2: AHORROS



CARACTERÍSTICA #1: EDAD

NOTE:  $P(\text{Non Retire} | X) = 1 - 0.25 = 0.75$



## MATRIZ DE CONFUSIÓN

