-Llamemos $p(C)$ a la probabilidad de que un visitante de un sitio de e-commerce **compre** luego de ver una publicidad/oferta.
-Llamemos $p(C|E)$ a la probabilidad de C dado que el comprador visitó otros sitios *E* previamente.

Va a ser muy difícil contar con los suficientes casos observados de visitantes que hayan comprado (***C***) y que hayan presentado la misma evidencia ***E*** (sitios visitados previamente). Es más, lo más probable es que esta colección de factores que conforman la evidencia E particular no se hayan presentado nunca anteriormente. Entonces deberemos encontrar la forma de obtener la P(*C/E*) considerando las piezas de evidencia por separado y combinándolas de alguna manera en una **E** general.

Teorema de Bayes

$p(A \wedge B) = p(A) \cdot p(B/A)$ $\quad\quad$ $p(A \wedge B) = p(B) \cdot p(A/B)$

$$\Rightarrow \boxed{p(B \mid A) = \frac{p(A \mid B) \cdot p(B)}{p(A)}} \quad \textbf{(1)}$$

La probabilidad de que se cumpla la hipótesis H dado la evidencia E es:

$$\Rightarrow p(H \mid E) = \frac{p(E \mid H) \cdot p(H)}{p(E)}$$

Clasificador Naive Bayes

Verosimilitud

P. priori

P. Posteriori

$$p(C = c \mid \mathbf{E}) = \frac{p(\mathbf{E} \mid C = c) \cdot p(C = c)}{p(\mathbf{E})} \qquad \textbf{(2)}$$

P. Evidencia

- $p(C = c)$ → es la **Probabilidad a priori.** Puede ser subjetiva, fruto de otra aplicación de Bayes o una estimación frecuentista.

- $p(\mathbf{E} \,/ C = c)$ → es la **verosimilitud,** o la probabilidad, de la evidencia **E** dado la clase C=c.

- $p(\mathbf{E})$ → es la **probabilidad de la evidencia**.

- $p(C = c \,/\mathbf{E})$ es la **probabilidad a POSTERIORI** (nuestro objetivo)

Consideremos al vector de evidencias $\mathbf{E}^T = (e_1, e_2, \cdots, e_k)$

$$p(\mathbf{E}|C{=}c) = P(\boldsymbol{e_1} \wedge \boldsymbol{e_2} \wedge \boldsymbol{e_3} \wedge \cdots \wedge \boldsymbol{e_k} / c) \qquad \textbf{(3)}$$

Analicemos la independencia en la probabilidad condicional. Sabíamos que $p(A \wedge B) = p(A) \cdot p(B/A)$. Actualicémoslo con el condicionante /c

$$p(A \wedge B/C) = p(A/C) \cdot p(B/A \wedge C)$$

¿Pero qué pasaría si asumimos que A y B son independientes cuando están condicionada por C? O sea que $P(B/A \wedge C) = P(B/C)$.

$$\Rightarrow \boxed{p(A \wedge B/C) = p(A/C) \cdot p(B/C)}$$

Aplicado a (3) sería:

$$p(\mathbf{E}|C{=}c) = P(\boldsymbol{e_1} \wedge \boldsymbol{e_2} \wedge \boldsymbol{e_3} \wedge \cdots \wedge \boldsymbol{e_k} / c) = P(\boldsymbol{e_1}/c)\, P(\boldsymbol{e_2}/c)\, P(\boldsymbol{e_3}/c) \cdots P(\boldsymbol{e}_k/c)\ ) \qquad \textbf{(4)}$$

Y ahora podré calcular $P(\boldsymbol{e}_i/c)$ como la <u>proporción de $\boldsymbol{e}_i$ en la clase $\boldsymbol{c}$</u>

Reemplazando (4) en (2) tendremos que:

$$p(c \mid \mathbf{E}) = \frac{p(e_1 \mid c) \cdot p(e_2 \mid c) \cdots p(e_k \mid c) \cdot p(c)}{p(\mathbf{E})} \qquad \textbf{(5)}$$

Por la regla de probabilidad total sabemos que:

$$\begin{aligned} p(\mathbf{E}) &= p(\mathbf{E} \wedge c_0) + p(\mathbf{E} \wedge c_1) \\ &= p(\mathbf{E} \mid c_0) \cdot p(c_0) + p(\mathbf{E} \mid c_1) \cdot p(c_1) \end{aligned} \qquad \textbf{(6)}$$

Por nuestra suposición de independencia de (4) podemos expresar a (6) de la siguiente manera:

$$\begin{aligned} p(\mathbf{E}) = \; & p(e_1 \mid c_0) \cdot p(e_2 \mid c_0) \cdots p(e_k \mid c_0) \cdot p(c_0) \\ & + p(e_1 \mid c_1) \cdot p(e_2 \mid c_1) \cdots p(e_k \mid c_1) \cdot p(c_1) \end{aligned} \qquad \textbf{(7)}$$

Y ahora reemplazando (7) en (5) llegamos a la fórmula del que nos permite calcular las probabilidades a posteriori fácilmente a partir de los datos de la muestra (*dataset*)

Clasificador Naive Bayes

$$p(c_0 \mid \mathbf{E}) = \frac{p(e_1 \mid c_0) \cdot p(e_2 \mid c_0) \cdots p(e_k \mid c_0) \cdot p(c_0)}{p(e_1 \mid c_0) \cdot p(e_2 \mid c_0) \cdots p(e_k \mid c_0) \cdot p(c_0) + p(e_1 \mid c_1) \cdot p(e_2 \mid c_1) \cdots p(e_k \mid c_1) \cdot p(c_1)}$$

Matriz de Confusión

Datos reales

Predicciones del Modelo

		True default status		
		No	Yes	Total
Predicciones del Modelo	No	9,644	252	9,896
	Yes	23	81	104
	Total	9,667	333	10,000

Exactitud (accuracy**): Acc = (VP + VN)/ total**

Memoria-Sensitividad : Recall = VP / (VP + FN)

Precisión: Precision = VP / (VP + FP)

Balance Precisión-Memoria (*Precision- Recall trade-Off*)

- **Memoria-Sensitividad (**Recall. Sensitivity)**: Recall = VP / (VP + FN)**
  Porcentaje de **REALMENTE** POSITIVOS (VP+FN) que el modelo clasificó correctamente como POSITIVOS (V**P**).

  OBJETIVO: mide qué tan bien se controlan los FALSOS NEGATIVOS (ej: cáncer)

  PROBLEMAS: El modelo puede engañarnos y maximizar esta Métrica si predice siempre positivo, o sea bajando el umbral (treshold).

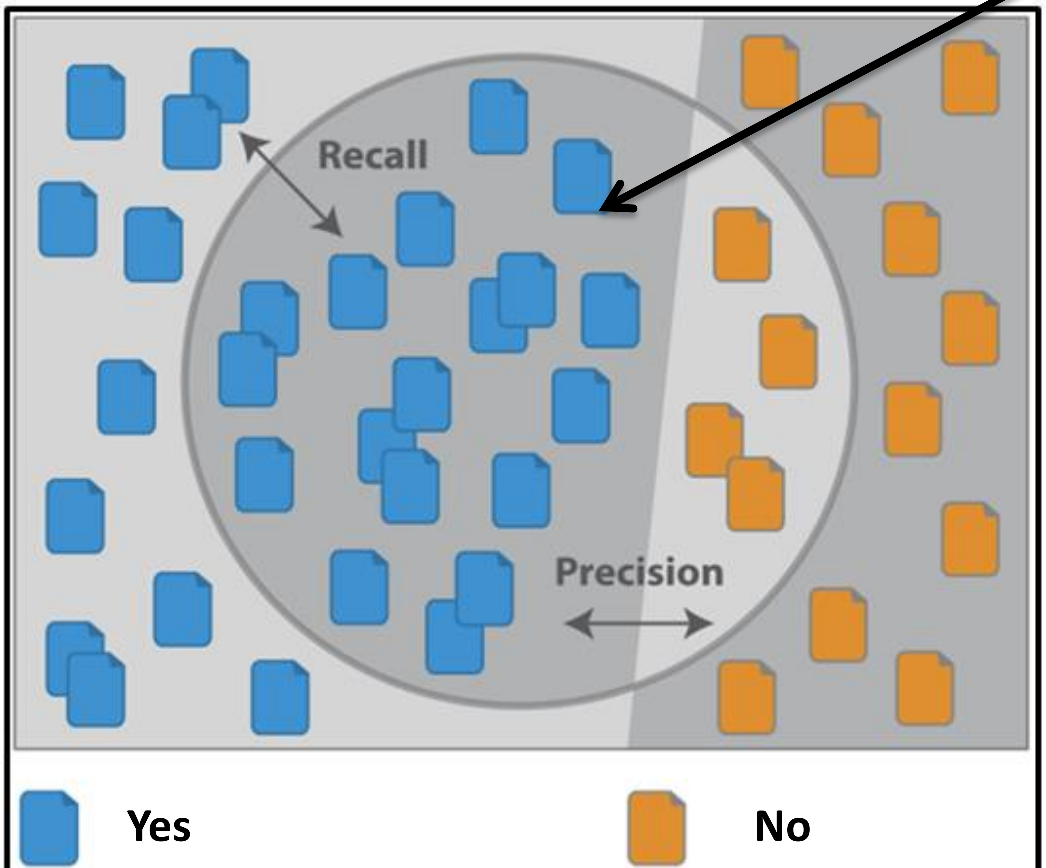
- **Precisión:** porcentaje de los **CLASIFICADOS** POSITIVOS (VP+FP)que realmente son POSITIVOS (**V**P). **Precision = VP / (VP + FP)**

  OBJETIVO: mide qué tan bien se controlan los FALSOS POSITIVOS (ej: Justicia)

  PROBLEMAS: El modelo puede engañarnos si sólo clasifica como positivos a los que tiene mucha confianza, o sea subiendo el umbral.

Matriz de Confusión

		Real		
		No	Yes	Total
Clasificado	No	9,644	252	9,896
	Yes	23	81	104
	Total	9,667	333	10,000



Matriz de Confusión

		Real		
		No	Yes	Total
Clasificado	No	9,644	252	9,896
	Yes	23	81	104
	Total	9,667	333	10,000