

Teorema de Bayes

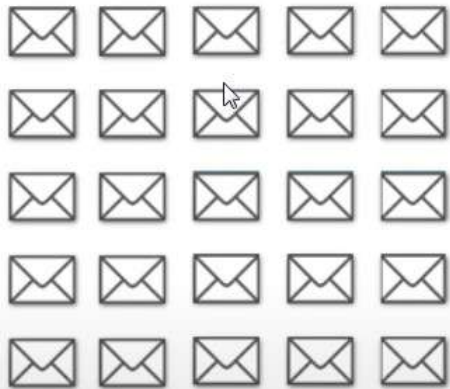
Detector de spam

100 e-mails



Detector de spam

25 Spam



75 No spam



Detector de spam



“Barato”

25 Spam



75 No spam



Detector de spam



“Barato”

25 Spam



75 No spam



Detector de spam



Spam

No spam



Quiz: Si un e-mail contiene la palabra "barato", cual es la probabilidad de que sea spam?

- ☐ 40%
- ☐ 60%
- ☐ 80%
- ☐ 100%

Detector de spam



“Barato”



80%

Spam



80%

No spam



20%

Quiz: Si un e-mail contiene la palabra “barato”, cual es la probabilidad de que sea spam?

- ☐ 40%
- ☐ 60%
- ☒ 80%
- ☐ 100%

Solución:
80%

Teorema de Bayes



“Barato”



80%

Spam

No spam



80%

20%

Quiz: Si un e-mail contiene la palabra “barato”, cual es la probabilidad de que sea spam?

- ☐ 40%
- ☐ 60%
- ☒ 80%
- ☐ 100%

Solución:
80%

Detector de spam

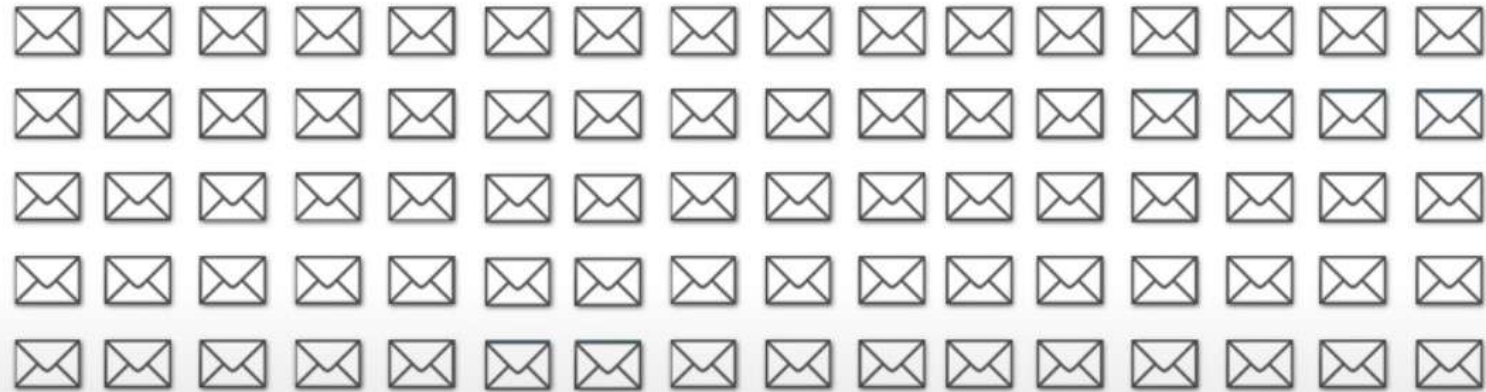


“Comprar”

Spam



No spam



Detector de spam



“Comprar”

Spam



No spam



Teorema de Bayes



“Comprar”

Spam



No spam



Quiz: Si un e-mail contiene la palabra “comprar”, cual es la probabilidad de que sea spam?

- ☐ 40%
- ☐ 60%
- ☐ 80%
- ☐ 100%

Teorema de Bayes



“Comprar”

Spam



60%

No spam



40%


Quiz: Si un e-mail contiene la palabra “comprar”, cual es la probabilidad de que sea spam?

Solución:
60%

- ☐ 40%
- ☒ 60%
- ☐ 80%
- ☐ 100%

Teorema de Bayes

 “Barato” \longrightarrow 80%

 “Comprar” \longrightarrow 60%

 “Barato” y “Comprar” \longrightarrow ?

Detector de spam



“Barato” y “Comprar”

Spam



No spam



Detector de spam



“Barato” y “Comprar”

Spam



No spam



Detector de spam

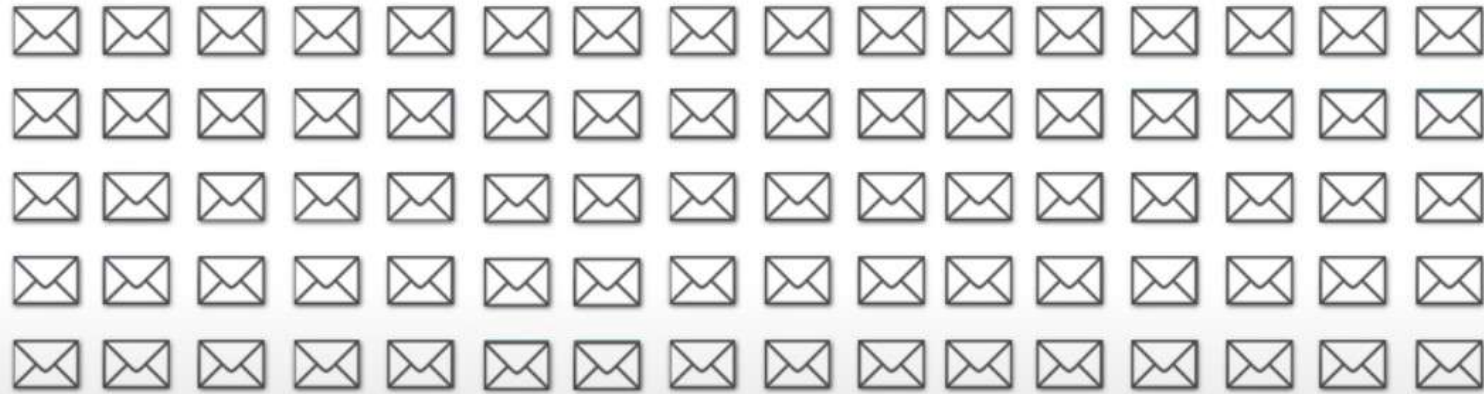


“Barato” y “Comprar”

Spam



No spam



Detector de spam



“Barato” y “Comprar”

Spam



No spam



Detector de spam



“Barato” y “Comprar”

Spam



No spam



Detector de spam



“Barato” y “Comprar”

Spam



No spam



Detector de spam



“Barato” y “Comprar”

Spam



No spam



Detector de spam



“Barato” y “Comprar”

Spam

No spam



Quiz: Si un e-mail contiene las palabras “barato” y “comprar”, cual es la probabilidad de que sea spam?

- ☐ 40%
- ☐ 60%
- ☐ 80%
- ☐ 100%

Detector de spam



“Barato” y “Comprar”



100% ?

Spam

No spam



100%



0%

Quiz: Si un e-mail contiene las palabras “barato” y “comprar”, cual es la probabilidad de que sea spam?

Solución:

100%

- ☐ 40%
- ☐ 60%
- ☐ 80%
- ☒ 100%

Problema



“Barato” y “Comprar”

Spam



12 e-mails

No spam



0 e-mails?

The background of the slide is a repeating pattern of small, stylized envelope icons. Each icon is a simple line drawing of an envelope with a triangular flap, arranged in a grid-like fashion across the entire background.

Solucion: Mas datos?

Solucion: Mas datos?

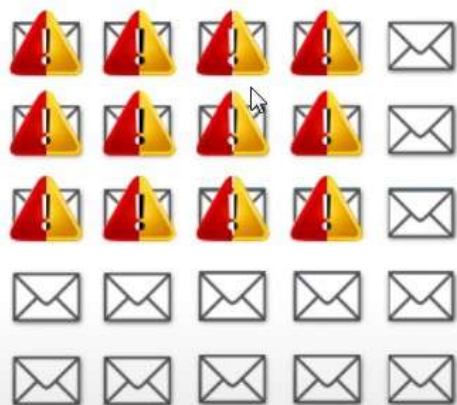


Detector de spam



“Barato” y “Comprar”

Spam



12 e-mails

No spam



0 e-mails?

Detector de spam



100 e-mails

Detector de spam



100 e-mails

5 “Barato”

10 “Comprar”

Detector de spam



100 e-mails

5 “Barato”

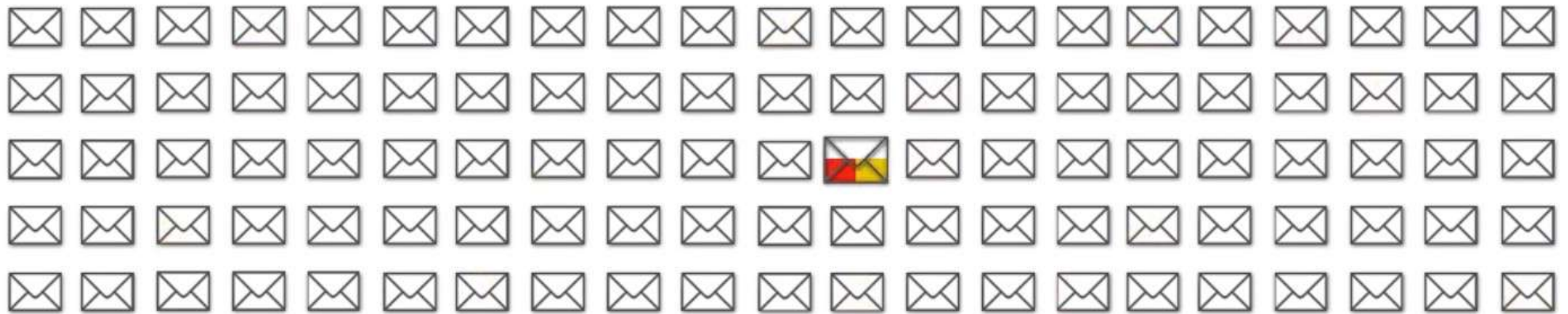
5% “Barato”

10 “Comprar”

10% “Comprar”

0.5% “Barato” y “Comprar”

Detector de spam



100 e-mails

5 "Barato"

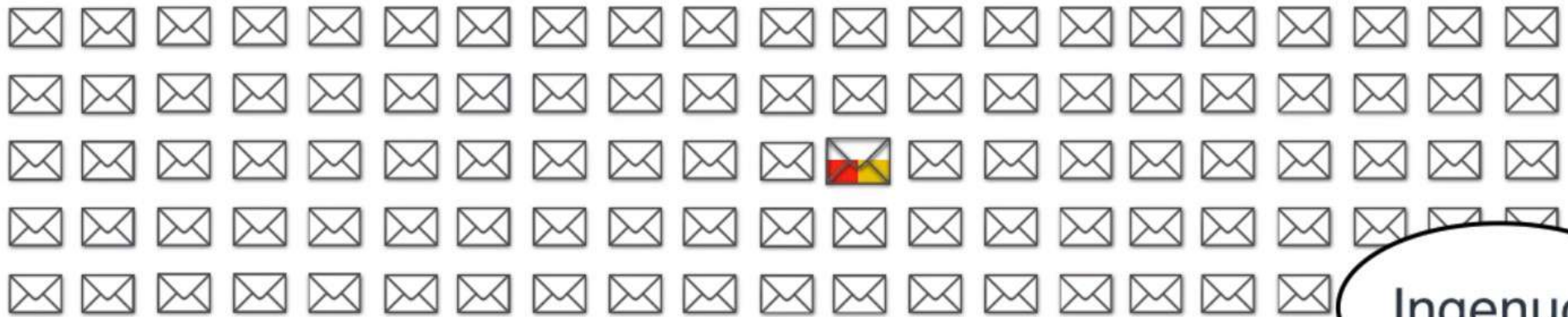
5% "Barato"

10 "Comprar"

10% "Comprar"

0.5% "Barato" y "Comprar"

Detector de spam



Ingenuo!

100 e-mails

5 "Barato"

10 "Comprar"

5% "Barato"

10% "Comprar"

Independientes

0.5% "Barato" y "Comprar"

Detector de spam

Spam



Detector de spam

Spam



25 e-mails

20 “Barato” 4/5

Detector de spam

Spam



25 e-mails

20 “Barato” 4/5

15 “Comprar” 3/5

Detector de spam

Spam



25 e-mails

20 "Barato"

15 "Comprar"

4/5

3/5

12/25

Detector de spam

Spam



25 e-mails

20 "Barato"

15 "Comprar"

4/5

3/5

→ 12/25

x 25

= 12 "Barato" and "Comprar"

Detector de spam

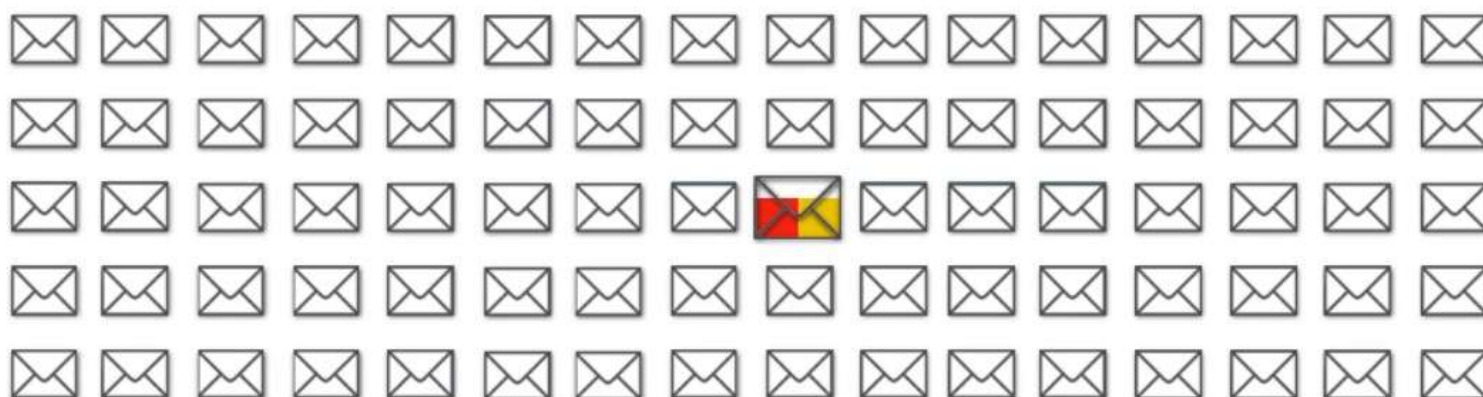
No spam



75 e-mails

Detector de spam

No spam



75 e-mails

5 "Barato"

1/15

10 "Comprar"

2/15

$$\begin{array}{l} 1/15 \\ 2/15 \end{array} \Rightarrow 2/225 \times 75 = 2/3 \text{ "Barato" y "Comprar"}$$

Detector de spam



“Barato” y “Comprar” → 94.737%

Spam

No spam

Quiz: Si un e-mail contiene las palabras “barato” y “comprar”, cual es la probabilidad de que sea spam?



12

$\underbrace{\hspace{1.5cm}}$
94.737%



2/3

$\underbrace{\hspace{1.5cm}}$
5.263%

$$\frac{12}{12 + 2/3} = \frac{36}{38} = 94.737\%$$

Detector de spam



“Barato” y “Comprar”

Spam



 12

No spam



 2/3

Bayes ingenuo

	Spam		No spam	
Total				
Barato				
Comprar				
Barato y Comprar				

Bayes ingenuo

	Spam		No spam	
Total	25		75	
Barato	20	4/5	5	1/15
Comprar	15	3/5	10	2/15
Barato y Comprar	12	12/25	2/3	2/225

$$\frac{12}{12 + 2/3} = \frac{36}{38} = 94.737\%$$

Bayes ingenuo

	Spam		No Spam	
Total	25		75	
Barato	20	4/5	5	1/15
Comprar	15	3/5	10	2/15
Trabajo	5	1/5	30	6/15
Barato, Comprar y Trabajo	12/5	12/125	4/15	12/3375

$$\frac{12/5}{12/5 + 4/15} = \frac{36}{40} = 90\%$$

Formulas?

Teorema de Bayes

S: Spam

H: Ham (no spam)

B: 'Barato'

$$P(S | B) = \frac{P(B | S) P(S)}{P(B | S) P(S) + P(B | H) P(H)}$$

Teorema de Bayes

S: Spam

H: Ham (no spam)

B: 'Barato'

$$P(S | B) = \frac{P(B | S) P(S)}{P(B | S) P(S) + P(B | H) P(H)}$$

$$P(\text{spam si "Barato"}) = \frac{\frac{20}{25} \frac{25}{100}}{\frac{20}{25} \frac{25}{100} + \frac{5}{75} \frac{75}{100}}$$

Teorema de Bayes

S: Spam

H: Ham (no spam)

B: 'Barato'

$$P(S | B) = \frac{P(B | S) P(S)}{P(B | S) P(S) + P(B | H) P(H)}$$

$$P(\text{spam si "Barato"}) = \frac{\frac{20}{25} \frac{25}{100}}{\frac{20}{25} \frac{25}{100} + \frac{5}{75} \frac{75}{100}} = 80\%$$

Teorema de Bayes

S: Spam

H: Ham (no spam)

C: 'Comprar'

$$P(S | C) = \frac{P(C | S) P(S)}{P(C | S) P(S) + P(C | H) P(H)}$$

$$P(\text{spam si "Comprar"}) = \frac{\frac{15}{25} \frac{25}{100}}{\frac{15}{25} \frac{25}{100} + \frac{10}{75} \frac{75}{100}} = 60\%$$

Bayes Ingenuo

$$P(\text{"Barato"} \ \& \ \text{"Comprar"}) = P(\text{"Barato"}) \ P(\text{"Comprar"})$$

$$P(B \cap C) = P(B) \ P(C)$$



Ingenuo

Bayes Ingenuo

S: Spam

H: Ham (not spam)

B: 'Barato'

C: 'Comprar'

$$P(S | B \cap C) = \frac{P(B \cap C | S) P(S)}{P(B \cap C | S) P(S) + P(B \cap C | H) P(H)}$$

Bayes Ingenuo

S: Spam

H: Ham (not spam)

B: 'Barato'

C: 'Comprar'

$$P(S | B \cap C) = \frac{P(B|S)P(C|S)P(S)}{P(B|S)P(C|S)P(S) + P(B|H)P(C|H)P(H)}$$

Bayes Ingenuo

S: Spam

H: Ham (not spam)

B: 'Barato'

C: 'Comprar'

$$P(S | B \cap C) = \frac{P(B|S)P(C|S)P(S)}{P(B|S)P(C|S)P(S) + P(B|H)P(C|H)P(H)}$$

$$P(\text{spam if "Barato" \& "Comprar"}) = \frac{\frac{20}{25} \cdot \frac{15}{25} \cdot \frac{25}{100}}{\frac{20}{25} \cdot \frac{15}{25} \cdot \frac{25}{100} + \frac{5}{75} \cdot \frac{10}{75} \cdot \frac{75}{100}} = 94.737\%$$