Spam Detector

100 e-mails

Spam Detector

25 Spam

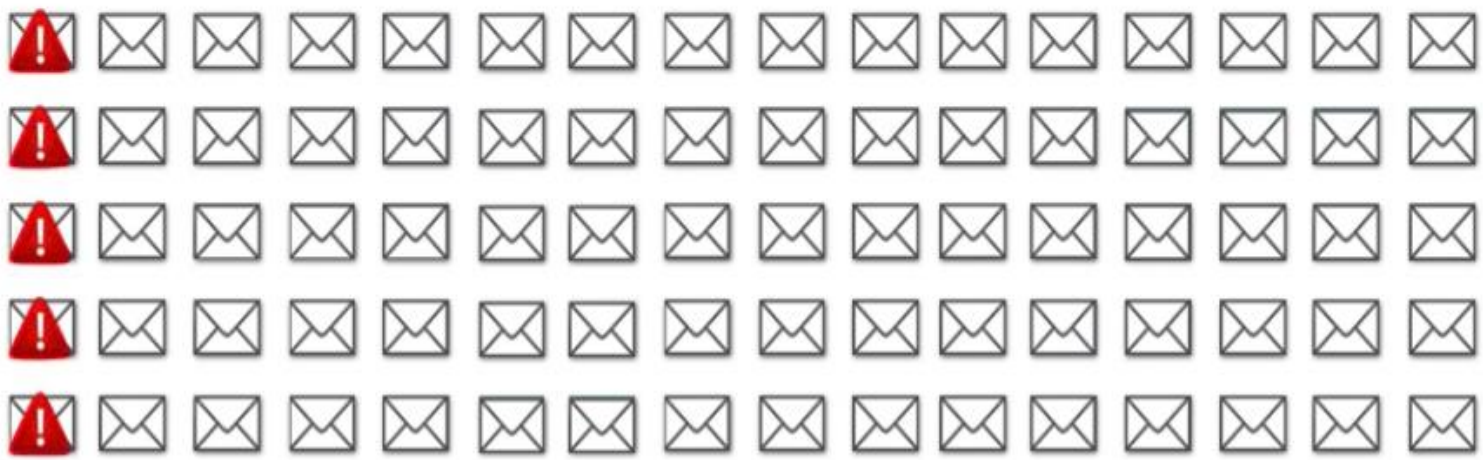
75 No spam

Spam Detector

"Buy"

25 Spam

75 No spam



Spam Detector

"Buy"

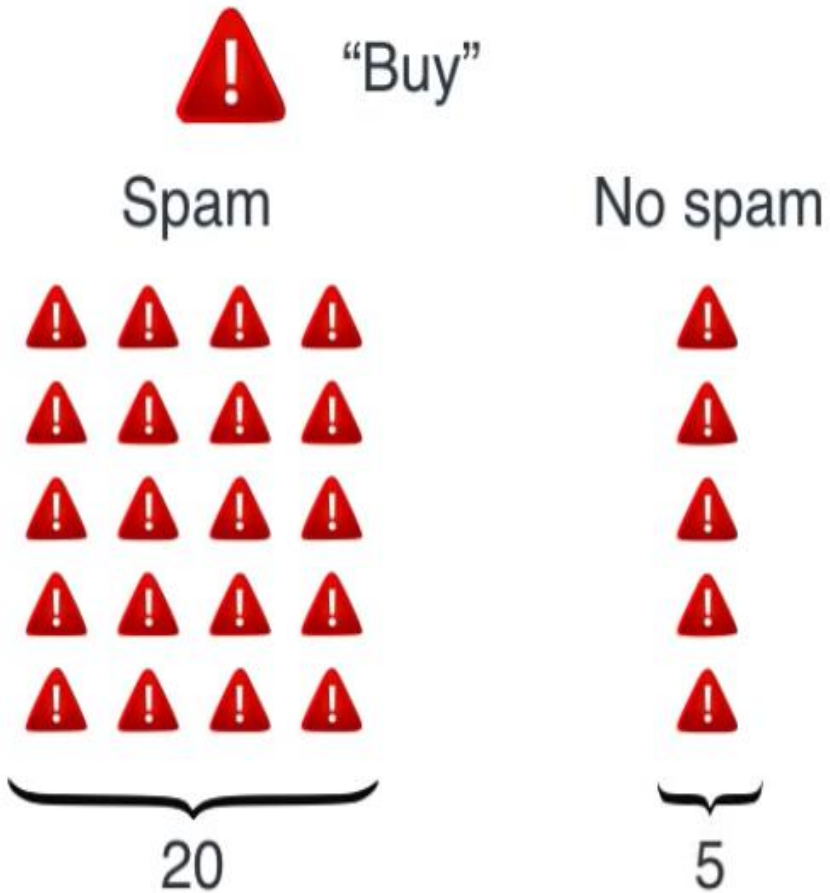
Spam

No spam

Quiz: If an e-mail contains the word "buy", what is the probability that it is spam?

- ○ 40%
- ○ 60%
- ○ 80%
- ○ 100%

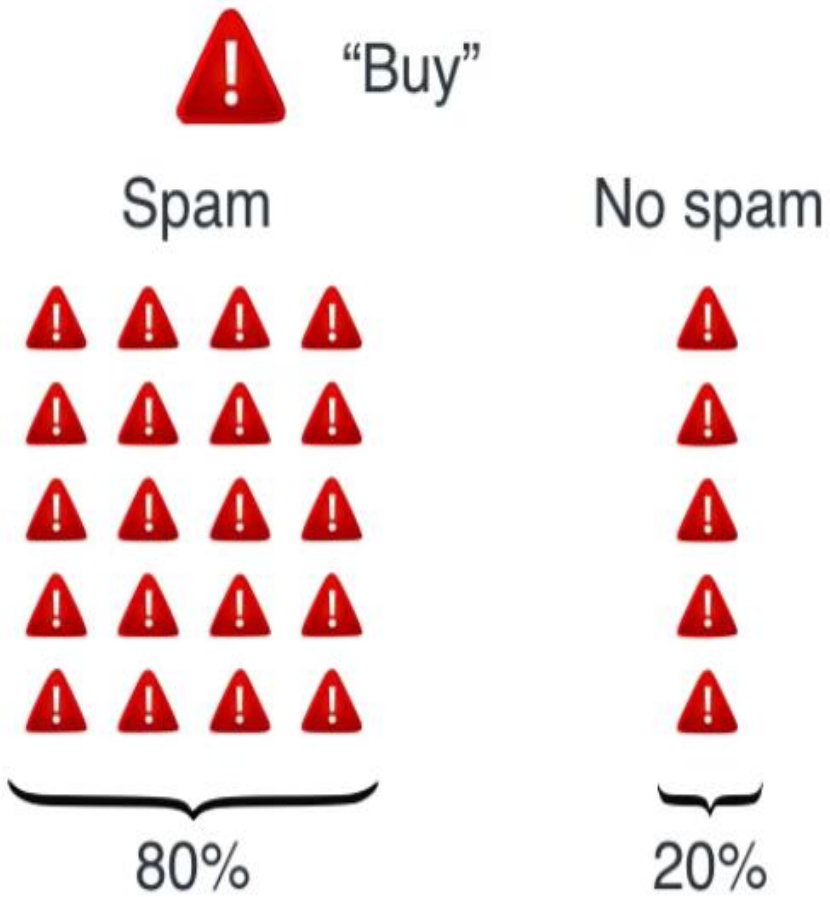
Spam Detector



Quiz: If an e-mail contains the word "buy", what is the probability that it is spam?

- ○ 40%
- ○ 60%
- ○ 80%
- ○ 100%

Spam Detector

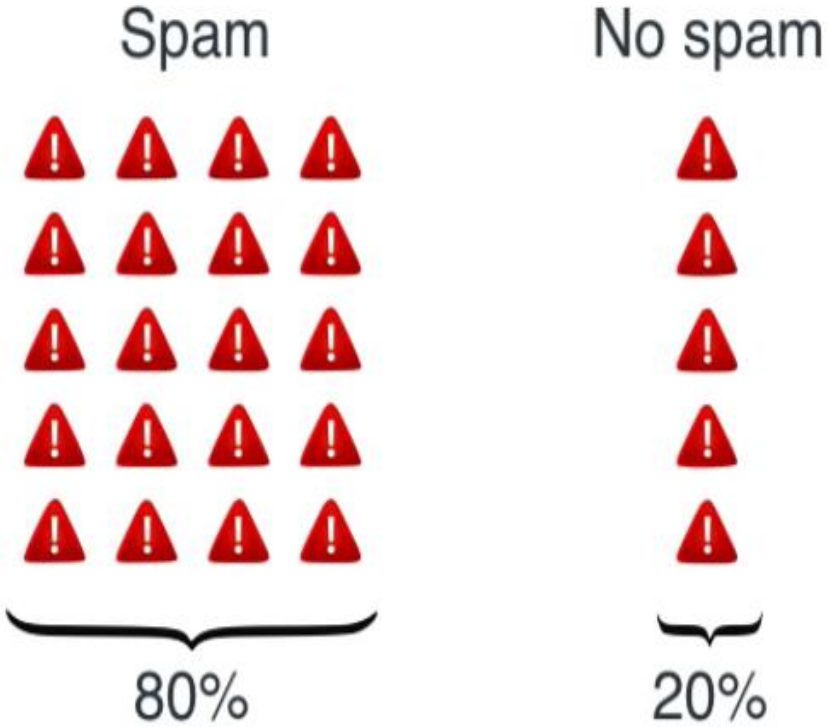


Quiz: If an e-mail contains the word "buy", what is the probability that it is spam?

- ○ 40%
- ○ 60%
- ○ 80%
- ○ 100%

Spam Detector

"Buy"



Quiz: If an e-mail contains the word "buy", what is the probability that it is spam?

- ○ 40%
- ○ 60%
- ● 80%
- ○ 100%

Solution:
80%

Bayes Theorem

 "Cheap"

Spam

No spam

Quiz: If an e-mail contains the word "cheap", what is the probability that it is spam?

- ○ 40%
- ○ 60%
- ○ 80%
- ○ 100%

Bayes Theorem

Quiz: If an e-mail contains the word "cheap", what is the probability that it is spam?

- ○ 40%
- ● 60%
- ○ 80%
- ○ 100%

Solution:
60%

Spam Detector

"Buy" and "Cheap"

Spam

No spam

Spam Detector

"Buy" and "Cheap"

Spam

No spam

Spam Detector

"Buy" and "Cheap"

Spam

No spam

Spam Detector

"Buy" and "Cheap"

Spam

No spam

Spam Detector

"Buy" and "Cheap"

Spam

100%

No spam

0%

Quiz: If an e-mail contains the words "buy" and "cheap", what is the probability that it is spam?

- ○ 40%
- ○ 60%
- ○ 80%
- ● 100%

Spam Detector

100 e-mails

5 “Buy”

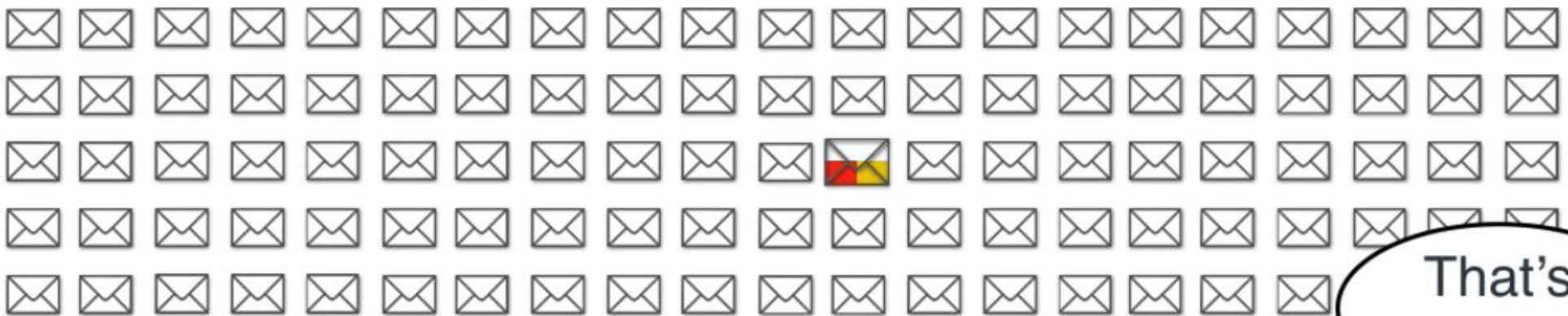
10 “Cheap”

5% “Buy”

10% “Cheap”

0.5% “Buy” and “Cheap”

Spam Detector



That's naive!

100 e-mails

5 "Buy"

10 "Cheap"

5% "Buy"

10% "Cheap"

Independent

0.5% "Buy" and "Cheap"

Naive Bayes Classifier

"Buy" and "Cheap" ⟶ 94.737%

Spam | No spam

 12 |  2/3

94.737% | 5.263%

Quiz: If an e-mail contains the words "buy" and "cheap", what is the probability that it is spam?

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

$$= 94.737\%$$

Naive Bayes

	Spam		No spam	
Total	25		75	
Buy	20	4/5	5	1/15
Cheap	15	3/5	10	2/15
Buy & Cheap	12	12/25	2/3	2/225

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

Bayes Theorem

S: Spam
H: Ham (not spam)
B: 'Buy'

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

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

Naive Bayes

P("Buy" & "Cheap") = P("Buy") P("Cheap")

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

Naive

Naive Bayes

S: Spam
H: Ham (not spam)
B: ‘Buy’
C: ‘Cheap’

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

$$P(\text{spam if “Buy” \& “Cheap”}) = \frac{\frac{20}{25}\,\frac{15}{25}\,\frac{25}{100}}{\frac{20}{25}\,\frac{15}{25}\,\frac{25}{100} + \frac{5}{75}\,\frac{10}{75}\,\frac{75}{100}}$$

$$= 94.737\%$$