

CHALLENGE PROJECT → BAYES

ALEXANDRA DE CELLA

CALCULATIONS:

* SPAM FILTERING

1) WORDS: "WATCH ANIME NOW"

$$o \text{ "WATCH"} \rightarrow P(\text{"WATCH"} | \text{SPAM}) \rightarrow 0.095$$

$$\hookrightarrow P(\text{"WATCH"} | \text{NOT SPAM}) \rightarrow 0.053$$

$$o \text{ "ANIME"} \rightarrow P(\text{"ANIME"} | \text{SPAM}) \rightarrow 0.095$$

$$\hookrightarrow P(\text{"ANIME"} | \text{NOT SPAM}) \rightarrow 0.053$$

$$o \text{ "NOW"} \rightarrow P(\text{"NOW"} | \text{SPAM}) \rightarrow \frac{1+1}{8+13} = 0.095$$

$$\hookrightarrow P(\text{"NOW"} | \text{NOT SPAM}) \rightarrow \frac{1}{6+13} = 0.053$$

$$* P(\text{"WATCH ANIME NOW"} | \text{SPAM}) = 3(0.095) \Rightarrow 0.285$$

$$* P(\text{"WATCH ANIME NOW"} | \text{NOT SPAM}) = 3(0.053) = 0.159$$

o o → "WATCH ANIME NOW" IS LIKELY SPAM

2) WORDS: "TAKE OUT AND ANIME AT MY HOUSE"

$$* \text{"TAKEOUT"} \rightarrow P(\text{"TAKEOUT"} | \text{SPAM}) = 0.0476$$

$$\hookrightarrow P(\text{"TAKEOUT"} | \text{NOT SPAM}) = 0.105$$

$$* \text{"AND"} \rightarrow P(\text{"AND"} | \text{SPAM}) = 0.0476$$

$$\hookrightarrow P(\text{"AND"} | \text{NOT SPAM}) = 0.053$$

$$* \text{"ANIME"} \rightarrow P(\text{"ANIME"} | \text{SPAM}) = 0.095$$

$$\hookrightarrow P(\text{"ANIME"} | \text{NOT SPAM}) = 0.053$$

$$* \text{"AT"} \rightarrow P(\text{"AT"} | \text{SPAM}) = 0.0476$$

$$\hookrightarrow P(\text{"AT"} | \text{NOT SPAM}) = 0.105$$

$$* \text{"MY"} \rightarrow P(\text{"MY"} | \text{SPAM}) = 0.0476$$

$$\hookrightarrow P(\text{"MY"} | \text{NOT SPAM}) = 0.053$$

$$* \text{"HOUSE"} \rightarrow P(\text{"HOUSE"} | \text{SPAM}) = 0.095$$

$$\hookrightarrow P(\text{"HOUSE"} | \text{NOT SPAM}) = 0.105$$

$$\circ P(\text{"TAKEOUT AND ANIME AT MY HOUSE"} | \text{SPAM}) = 4(0.0476) + 2(0.095) = 0.38$$

$$\circ P(\text{"TAKEOUT AND ANIME AT MY HOUSE"} | \text{NOT SPAM}) = 3(0.105) + 3(0.053) = 0.474$$

$\circ \circ$ ("TAKEOUT AND ANIME AT MY HOUSE") \rightarrow MORE LIKELY NOT SPAM

3) WORDS: "SELL ME YOUR ANIME COLLECTION"

$$* \text{"SELL"} \rightarrow P(\text{"SELL"} | \text{SPAM}) \rightarrow 0.095$$

$$\hookrightarrow P(\text{"SELL"} | \text{NOT SPAM}) \rightarrow 0.105$$

$$* \text{"ME"} \rightarrow P(\text{"ME"} | \text{SPAM}) \rightarrow 0.0476$$

$$\hookrightarrow P(\text{"ME"} | \text{NOT SPAM}) \rightarrow 0.053$$

$$* \text{"YOUR"} \rightarrow P(\text{"YOUR"} | \text{SPAM}) \rightarrow 0.095$$

$$\hookrightarrow P(\text{"YOUR"} | \text{NOT SPAM}) \rightarrow 0.053$$

$$* \text{"ANIME"} \rightarrow P(\text{"ANIME"} | \text{SPAM}) \rightarrow 0.095$$

$$\hookrightarrow P(\text{"ANIME"} | \text{NOT SPAM}) \rightarrow 0.053$$

$$* \text{"COLLECTION"} \rightarrow P(\text{"COLLECTION"} | \text{SPAM}) \rightarrow 0.0476$$

$$\hookrightarrow P(\text{"COLLECTION"} | \text{NOT SPAM}) \rightarrow 0.053$$

$$o P(\text{"SELL ME YOUR"} \overset{\text{ANIME}}{\downarrow} \text{"COLLECTION"} | \text{SPAM}) = 3(0.095) + 2(0.0476) = 0.38$$

$$o P(\text{"SELL ME YOUR"} \overset{\text{ANIME}}{\downarrow} \text{"COLLECTION"} | \text{NOT SPAM}) = 0.105 + 3(0.053) = 0.264$$

o "SELL ME YOUR ~~YOUR~~ ANIME COLLECTION" MORE LIKELY SPAM.