

# SPAM DETECTOR

ตัวตรวจจับ **spam** หรือ **spam classifier** อัลกอริทึมที่จะใช้สร้างจะมีความเกี่ยวข้องกับทฤษฎีทางคณิตศาสตร์ที่เรียกว่า ทฤษฎีบทของเบย์ส

ความน่าจะเป็นแบบมีเงื่อนไข

**$P(A|B)$**  ถูกกำหนดให้เป็นความน่าจะเป็นที่จะเกิดเหตุการณ์ **A** ขึ้น ถ้าเกิด  
มีเหตุการณ์ **B** เกิดขึ้น ซึ่งจะสามารถคำนวณได้จาก ความน่าจะเป็นที่จะเกิด  
เหตุการณ์ **A** และ **B** พร้อมกันหารกับความน่าจะเป็นที่จะเกิดเหตุการณ์ **B**

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

Probability of event A  
given B has occurred

Probability of event A occurred  
and event B occurred

Probability of event B

ความน่าจะเป็นแบบมีเงื่อนไข

ทฤษฎีบทของเบย์ส ทำให้สามารถคำนวณความน่าจะเป็นที่จะเกิดเหตุการณ์หนึ่ง โดยอิงมาจากอีกเหตุการณ์หนึ่งได้  
ปกติแล้วจะใช้คำนวณหาความน่าจะเป็นของเหตุการณ์ที่วัดได้ยาก โดยจะคำนวณได้จากเหตุการณ์อีกเหตุการณ์ที่เกี่ยวข้องกันแทน ทำให้สามารถคำนวณได้ว่า ความเป็นไปได้ที่ประโยชน์นั้นเป็นสแปมจะเท่ากับ พจน์ที่เหลือ

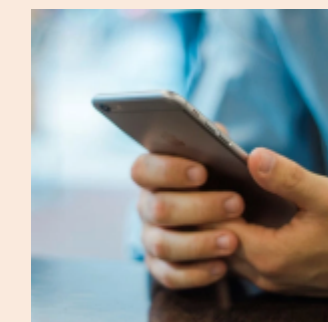
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$P(\text{spam}|\text{penis}) = \frac{P(\text{penis}|\text{spam}) * P(\text{spam})}{P(\text{penis})}$$

$$= \frac{\frac{20}{30} * \frac{30}{74}}{\frac{51}{74}} = \frac{20}{51} = 0.39$$

# DATA SET

▲ v1		▲ v2	
class		sms	
ham spam	87% 13%	5169 unique values	
ham		Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got a...	
ham		Ok lar... Joking wif u oni...	
spam		Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entr...	
ham		U dun say so early hor... U c already then say...	
ham		Nah I don't think he goes to usf, he	



## SMS Spam Collection Dataset

Collection of SMS messages tagged as spam or legitimate

[k](https://www.kaggle.com) kaggle.com

**THANK**