Assignment II

⇒ Given

Not Spam.	Spam.
① Hi there, how are you today?	① Win a free price now!
② Meeting at 3PM today!	② clain your discount tod
③ Please send the Report.	③ limited time offer. click here,

now.

$P(\text{Not Spam}) = 3/6 = 0.5$

$P(\text{Spam}) = 3/6 = 0.5$

Total vocabulary Size (V) = 28.

notspam word (N_{NS}) = 14

Spam word (N_S) = 14.

Classifying "free meeting tomorrow"

for not spam.

$P(\text{Free}|\text{Notspam}) = 0+1/42$

$P(\text{Meeting}|\text{notspam}) = 1+1/42$

$P(\text{tommorow}|\text{notspam}) = 1+1/42$

$P(\text{Total}|\text{Notspam}) = P(N_S) \times [P(\text{Free}|\text{Notspam}) \times P(\text{meeting}|\text{notspam}) \times P(\text{tomorro}$

. |not spam)

$= 0.5 \times 1/42 + 2/42 + 2/42$

$= 2/74,088 = 0.00002699$

for Spam.

$P(\text{free} \mid \text{spam}) = 1+1/42 = 2/42$.

$P(\text{meeting} \mid \text{spam}) = 0+1/42 = 1/42$.

$P(\text{tomorrow} \mid \text{spam}) = 0+1/42 = 1/42$

$P(\text{Total} \mid \text{spam}) = P(s) \times P(\text{free} \mid \text{spam}) \times P(\text{meeting} \mid \text{spam}) \times P(\text{tomorrow} \mid \text{spam})$

$= 0.5 \times 2/42 + 1/42 + 1/42$

$= 0.000013497$.

now

for normalization. of "free meeting tomorrow"

$$P(NS \mid N) = \frac{P(\text{Total} \mid \text{Not Spam})}{P(\text{total} \mid \text{Not Spam}) + P(\text{total} \mid \text{Spam})}$$

$$= \frac{0.00002699}{0.00002699 + 0.000013497} \times 100\%$$

$= 66.67\%$

$$P(S \mid N) = \frac{P(\text{Total} \mid \text{Spam})}{P(\text{Total} \mid \text{spam}) + P(\text{total} \mid \text{Not Spam})}$$

$$= \frac{0.000013497}{0.00002699 + 0.000013497} \times 100\%$$

$= 33.33\%$.

Since $P(NS \mid N) > P(S \mid N)$ "free meeting tomorrow" is not Spam.

Again
classifying "claim your free price"

Since All words ("claim", "your", "free", "price") appeared 0 times in Not Spam.

Probability = 0 + 1/42 = 1/42.

$P(\text{Total} \mid \text{NotSpam}) = 0.5 \times 1/42 \times 1/42 \times 1/42 \times 1/42$
$= 0.5/3111,696$

again
for Spam.
Probability = 1 + 1/42 = 2/42

$P(\text{Total} \mid \text{Spam}) = 0.5 \times 2/42 \times 2/42 \times 2/42 \times 2/42$
$= 8/3,111,696$

for Normalization.

Total Sum = 0.5 + 8 (ignoring common denominator)
= 8.5.

$P(NS \mid N) = 0.5/8.5 \approx 5.88\%$.
$P(S \mid N) = 8/8.5 \approx 94.12\%$.

Since $P(S \mid N) > P(NS \mid N)$ its Spam