1. Using the Naive Bayes Algorithm, how would you classify a new email that contained both the words ‘viagra’ and ‘unsubscribe’?
2. Using the Naive Bayes Algorithm, how would you classify a new email that doesn’t contain either 'viagra' or unsubscribe?

Frequency Table:

	Viagra	Not Viagra	Unsubscribe	Not unsubscribe	
Spam	4	36	11	29	40
Ham	1	159	10	150	160
	5	195	21	179	200

S = Spam
H = Ham
V = Viagra
U = Unsubscribe

Probabilities:

$P(S) = 40/200$
$P(H) = 160/200$
$P(V|S) = 4/40$
$P(\sim V|S) = 36/40$
$P(U|S) = 11/40$
$P(\sim U|S) = 29/40$
$P(V|H) = 1/160$
$P(\sim V|H) = 159/160$
$P(U|H) = 10/160$
$P(\sim U|H) = 150/160$

1. Using the Naive Bayes Algorithm, how would you classify a new email that contained both the words ‘viagra’ and ‘unsubscribe’?

P(S | V, U) = P(V,U | S) * P(S)/ P(V,U)

Law of total probabilities:

P(V, U) = P(V,U | S) * P(S)+ P(V,U | H) * P(H)

Plugging in all the values into:

P(S | V, U) = P(V,U | S) * P(S)/ [P(V,U | S) * P(S)+ P(V,U | H) * P(H)]

P(V, U | S) = P(V | S) * P(U | S) due to class conditional independence

P(V, U | H) = P(V | H) * P(U | H)

= 40/200 * 4/40 * 11/40 / (40/200 * 4/40 * 11/40 + 160/200 * 1/160 * 10/160)

= 0.9462

There is a 94.62% chance of an email being spam if it contains both the words Viagra and Unsubscribe.

2. Using the Naive Bayes Algorithm, how would you classify a new email that doesn’t contain either 'viagra' or unsubscribe?

   P(S | ~V, ~U) = P(~V,~U | S) * P(S)/ P(~V,~U)

   Law of total probabilities:

   P(~V, ~U) = P(~V,~U | S) * P(S)+ P(~V,~U | H) * P(H)

   Plugging in all the values into:

   P(S | ~V, ~U) = P(~V,~U | S) * P(S)/ [P(~V,~U | S) * P(S)+ P(~V,~U | H) * P(H)]

   P(~V, ~U | S) = P(~V | S) * P(~U | S) due to class conditional independence

   P(~V,~U | H) = P(~V | H) * P(~U | H)

   = 36/40 * 29/40 * 40/200 / (36/40 * 29/40 * 40/200 + 160/200 * 159/160* 150/160)

   = 0.14900

   There is a 14.9% chance of an email being spam if it contains none of the words Viagra and Unsubscribe.