1) Given that the local pub is crowded, what is the probablity that Manchester united won?

→ M:- Manchester united wins the game.
P:- The pub is packed.

$P(M) = 0.7$

$P(!M) = 1 - P(M) = 1 - 0.7 = 0.3$

$P(P|M) = 0.9$

$P(P|!M) = 0.6$

∴ we need to calculate $P(M|P)$.

$$\therefore P(M|P) = \frac{P(P|M)\,P(M)}{P(P)}$$

$$= \frac{P(P|M)\,P(M)}{P(P|M)P(M) + P(P|!M)P(!M)}$$

$$= \frac{(0.9 \times 0.7)}{(0.9 \times 0.7) + (0.6 \times 0.3)}$$

$$= \frac{0.63}{0.63 + 0.18}$$

$$= 0.63/0.81$$

$$= \frac{7}{9} = 0.778$$

∴ After observing the probablity of 0.778, we can assume that Manchester united gonna win.

2) Given that Mr. Smith died, what is the probablity the nurse forgot to give him the pill?

→ F : The nurse forgets to give the pill.

D : Mr. Smith dies.

$P(F) = 0.3$

$P(!F) = 1 - 0.3 = 0.7$

$P(D|!F) = 0.1$

$P(D|F) = 0.8$

we need to calculate $P(F|D)$

$$\therefore P(F|D) = \frac{P(D|F)\,P(F)}{P(D)}$$

$$= \frac{P(D|F)\,P(F)}{P(D|F)P(F) + P(D|!F)P(!F)}$$

$$= \frac{(0.8 \times 0.3)}{(0.8)(0.3) + (0.1)(0.7)}$$

$$= \frac{0.24}{0.24 + 0.07}$$

$$= \frac{0.24}{0.31} = 0.774.$$

∴ Given Mr. Smith died, we say there is 77% chance the nurse forgot to give pill.

3) what is the probablity of gold being found on campus

⇒ G: Gold in campus.

c: coal in campus.

N: No mineral in campus.

p: The test give positive result.

$p(G) = 0.1$ | $p(p|G) = 0.8$

$p(c) = 0.3$ | $p(p|c) = 0.4$

$p(N) = 0.6$ | $p(p|N) = 0.2$

we have to calculate $p(G|p)$

$$\therefore p(G|p) = \frac{p(p|G)\,p(G)}{p(p)}$$

$$= \frac{p(p|G)\,p(G)}{p(p|G)\,p(G) + p(p|c)\,p(c) + p(p|N)\,p(N)}$$

$$= \frac{(0.8 \times 0.1)}{(0.8 \times 0.1) + (0.4 \times 0.3) + (0.2 \times 0.6)}$$

$$= \frac{0.08}{0.08 + 0.12 + 0.12}$$

$$= \frac{0.08}{0.32} = \frac{1}{4} = 0.25$$

$\therefore$ Given the Test give positive result, The chance of gold being found is 25%.

4) Given Positive test result, probablity of having Meningitis? A second Negative result effect the chance?

→ m: Patient has Meningits.

p: The test result is negative.

$p(m) = 0.05$

$p(p|m) = 0.95$

$p(p|!m) = 0.70$

(1) ∵ we have to calculate $p(m|p)$

$$\therefore p(m|p) = \frac{p(p|m)\,p(m)}{p(p)}$$

$$= \frac{p(p|m)\,p(m)}{p(p|m)\,(p(m) + p(p|!m)\,p(!m)}$$

$$= \frac{(0.95 \times 0.05)}{(0.95 \times 0.05) + (0.30)(0.95)}$$

$$= \frac{0.0475}{0.0475 + 0.2850}$$

$$= \frac{0.0475 \quad \cancel{95} \; 19}{0.3325 \quad \cancel{665} \; 113}$$

$$= \frac{19}{113} = 0.142$$

∴ Given a positive result, the chance of having meningitis is 14.2%.

(2) Lets say, $p1$ = Test one give positive result,
$p2$ = Test two give negative result.

$\therefore p(m|p1) = 0.142$

$p(!p2|m) = 1 - p(p1|m) = 1 - 0.95 = 0.05$

$p(!p2|!m) = 0.70$

$\therefore$ we need to calculate $p(m|p1, !p2)$

$$\therefore p(m|p1, !p2) = \frac{p(!p2|m)\, p(m|p1)}{p(!p2)}$$

$$= \frac{p(!p2|m)\, p(m|p1)}{p(!p2|m)p(m|p1) + p(!p2|!m)p(!m|p1)}$$

$$= \frac{(0.05 \times 0.1429)}{(0.05 \times 0.1429) + (0.70 \times 0.857)}$$

$$= \frac{0.007145}{0.007145 + 0.60007}$$

$$= \frac{0.007145}{0.6072} = 0.0118$$

$\therefore$ After, secon Test (negative) the probablity become 1.18% from 14.29%, means its significently reduce the chane of having meningites.

⊙ Seeing the probablity changes (5% → 14.2% → 1.18%) we can say the pacient don't have meningitis.

5) Given a negative test, can we say it's not gonna rain?

⇒ R : Rain on the day.
N : Test result Negative.
P : Test result positive.

∴ $P(R) = 0.8$
$P(P|R) = 0.75$
$P(P|!R) = 0.15$

∴ $P(N|R) = 1 - P(P|R) = 0.25$
∴ $P(N|!R) = 1 - P(P|R) = 0.85$

∴ we need to calculate $P(!R|N)$

$$\therefore P(!R|N) = \frac{P(N|!R)\,P(!R)}{P(N)}$$

$$= \frac{P(N|!R)\,P(!R)}{P(N|R)\,P(R) + P(N|!R)\,P(!R)}$$

$$= \frac{(0.85 \times 0.2)}{(0.25 \times 0.8) + (0.85 \times 0.2)}$$

$$= \frac{0.17}{0.2 + 0.17}$$

$$= \frac{17}{37} = 0.459$$

∴ Given a negative test, there is 45% chance of not rain. So 55% chance of raining.

Therefore Mr. Jons should not predict for no rain for that day.

6) calculate the probablity of having chickunguniya.

→ C: Fred has chickunguya
J: Fred complains joint pain
T: positive test result.

$P(C) = 0.0001$
$P(T|C) = 0.99$
$P(T|!C) = 0.04$

$P(J|C) = 0.64$
$P(J|!C) = 0.60$

1) we need to find $P(C|J, T_1)$

$$P(C|J) = \frac{P(J|C)\,P(C)}{P(J|C)P(C) + P(J|!C)\,P(!C)}$$

$$= \frac{(0.64)(0.0001)}{(0.64 \times 0.0001) + (0.60 \times 0.9999)} = \frac{0.000064}{0.600004}$$

$$P(C|J, T_1) = \frac{P(T|C) \times P(C|J)}{P(T|C)\,P(C|J) + P(T|!C)\,P(!C|J)}$$

$$= \frac{0.99 \times 0.0001067}{(0.99 \times 0.0001067) + (0.04 \times 0.9998933)}$$

$$= \frac{0.0001056}{0.0401013} = 0.00263$$

∴ seeing the probablity of 0.263%. we can say even he has join pain ~~there~~ he don't have chickunguniya.

(2) Second Test effect on result (T2)

$$p(C|S, T_1, T_2) = \frac{p(T|C)\,p(C|S, T_1)}{p(T|C)\,p(C|S, T_1) + p(T|!C)\,p(!C|S, T_1)}$$

$$= \frac{(0.99 \times 0.00263)}{(0.99 \times 0.00263) + (0.04 \times 0.99737)}$$

$$= \frac{0.002604}{0.042499}$$

$$= 0.0613$$

∴ A second positive test increses the probablity from 0.263% → 6.13%, but it is still low to say that he had chickungunya.

(3) Number of Tests need to confirm of desises.

→ Prior odds (after joint pain): $O(C|S) = \frac{P(C|S)}{P(!C|S)} = 0.0001067$

∴ Likehood Ratio: $LR(+) = \frac{P(T|C)}{P(T|!C)} = \frac{0.99}{0.04} = 24.75$

we need to find n such that:

$$O(C|S, T_1, \ldots T_n) = O(C|S) \times (LR(+))^n > 1$$

$$= 0.0001067 \times (24.75)^n > 1$$

Taking the logarithm:

$n(3.2087) > 9.1456$

$n > \frac{9.1456}{3.2087} = 2.85 \simeq 3$

∴ Essimated Test number is 3.

7. Probablity of a customer with specific attributes churning. Effect of 'm' on probablity when using m-estimate.

→ Given; Handset = (old), Time since customer = (>2.5), age = (≤55)

∴ Total customers = 10

Churned_yes = 5 ; Not_churned = 5

. P(yes) = 0.5   p(no) = 0.5

∴ we need to calculate,

① • Probablity of churning ⇒ $p(yes) \times p(old|yes) \times p(>2.5|yes) \times p(\leq 55|yes)$

$$= 0.5 \times \frac{2}{5} \times \frac{2}{5} \times \frac{3}{5}$$

$$= \frac{\cancel{5}}{10} \times \frac{\cancel{2}}{5} \times \frac{2}{5} \times \frac{3}{5}$$

$$= \frac{6}{125} = 0.048$$

• probablity of not churning ⇒

$$p(no) \times p(old|no) \times p(>2.5|no) \times p(\leq 55|no)$$

$$= 0.5 \times \frac{0}{5} \times \frac{3}{5} \times \frac{2}{5} = 0$$

∴ As the probablity of not_churning is 0, so the normalized probablity of customer churning is 100%.

(2) m-estimate formula is:-

$$p(x_i|y) = \frac{n_c + (m \times p)}{(n+m)}$$

n_c : number of feature
n : total count of class =
m ; parameter =
p = probablity of feature

8/2

For Churned = Yes

$\therefore p(old|yes) = \frac{2+1\times 0.5}{5+1} = \frac{2.5}{6}$

$\therefore p(>2.5|yes) = \frac{2+1\times 0.33}{5+1} = \frac{2.33}{6}$

$\therefore p(\leq 55|yes) = \frac{3+1\times 0.5}{5+1} = \frac{3.5}{6}$

For churned = No

$p(old|no) = \frac{0+1\times 0.5}{5+1} = \frac{0.5}{6}$

$\therefore p(>2.5|no) = \frac{3+1\times 0.33}{5+1} = \frac{3.33}{6}$

$\therefore p(\leq 55|no) = \frac{2+1\times 0.5}{5+1} = \frac{2.5}{6}$

$\therefore$ As the dataset is small, Lets essume $M=1$;

$p(yes) = 0.5 \times 0.417 \times 0.389 \times 0.583 = 0.047$

$p(no) = 0.5 \times 0.083 \times 0.556 \times 0.417 = 0.0096$

$\therefore$ Normalized churning : $\frac{0.0472}{0.0472+0.0096} = 0.831$

$\therefore$ As the given dataset is small, and more realiable than assumation, thats why a small m is more relavent than a big m. This prevents the posterior probility for an entire class from zero, which would cause the naive Bayes model to fail.

8) Calculate the probablity of customer being a positive example

- There is two output class: positive (+) & Negative (−)
- Total data point: 8 | positive count: 4 | Negative count: 4

Standard deviation (σ):- $\sqrt{\frac{1}{n-1}\sum(x_i-\mu)^2}$

For positive class (+):

x values: 5, 5, 3, 4

$$\mu_x^+ = \frac{5+5+3+4}{4} = \frac{17}{4}$$
$$= 4.25$$

$$\sigma_x^+ = \sqrt{\frac{(5-4.25)^2+(5-4.25)^2+(3-4.25)^2+(4-4.25)^2}{4-1}}$$

$$= \sqrt{\frac{0.5625+0.5625+1.5625+0.0625}{3}}$$

$$= \sqrt{\frac{2.75}{3}} = 0.957$$

y values: 6, 7, 6, 5

$$\mu_y^+ = \frac{6+7+6+5}{4} = \frac{24}{4} = 6$$

$$\sigma_y^+ = \sqrt{\frac{(6-6)^2+(7-6)^2+(6-6)^2+(5-6)^2}{4-1}}$$

$$= \sqrt{\frac{0+1+0+1}{3}}$$

$$= \sqrt{\frac{2}{3}} = 0.816$$

For Negative class(-):

x values: 1, 3, 2, 8

$$\mu_x^- = \frac{1+3+2+8}{4} = \frac{14}{4} = 3.5$$

$$\sigma_x^- = \sqrt{\frac{(1-3.5)^2+(3-3.5)^2+(2-3.5)^2+(8-3.5)^2}{4-1}}$$

$$= \sqrt{\frac{6.25+0.25+2.25+20.25}{3}}$$

$$= \sqrt{\frac{29}{3}} = 3.109$$

y values: 7, 4, 6, 1

$$\mu_y^- = \frac{7+4+6+1}{4} = \frac{18}{4} = 4.5$$

$$\sigma_y^- = \sqrt{\frac{(7-4.5)^2+(4-4.5)^2+(6-4.5)^2+(1-4.5)^2}{4-1}}$$

$$= \sqrt{\frac{6.25+0.25+2.25+12.25}{3}}$$

$$= \sqrt{\frac{21}{3}} = \sqrt{7} = 2.646$$

Given data for customer; $x = 7$, $y = 4$

Gaussian density function $N(x : \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}}$

Likelihood for the positive class(+):

$$p(x=7|+) = N(7; 4.25, 0.957)$$

$$= \frac{1}{\sqrt{2\pi}(0.957)} e^{-\frac{(7-4.25)^2}{2(0.957)^2}}$$

$$= \frac{1}{2.40} e^{-4.07} = 0.0101$$

$$p(y=4|+) = N(4; 6, 0.816).$$

$$= \frac{1}{\sqrt{2\pi}(0.816)} e^{-\frac{(4-6)^2}{2(0.816)^2}}$$

$$= \frac{1}{2.04} e^{-2.99} = 0.0163$$

Likelihood for the Negative class(-):

$$p(x=7|-) = N(7; 3.5, 3.109)$$

$$= \frac{1}{\sqrt{2\pi}(3.109)} e^{-\frac{(7-3.5)^2}{2(3.109)^2}}$$

$$= \frac{1}{7.80} e^{-0.64} = 0.0463$$

$$p(y=4|-) = N(4; 4.5, 2.646)$$

$$= \frac{1}{\sqrt{2\pi}(3.109)} e^{-\frac{(7-3.5)^2}{(2.3109)^2}}$$

$$= \frac{1}{7.80} e^{-0.45} = 0.1465$$

unnormalized Posterior for positive:

$p(+) \times p(x=7|+) \times p(y=4|+)$

$= 0.5 \times 0.0101 \times 0.0163 = 0.0000822$

unnormalized posterior for Negative;

$p(-) \times p(x=7|-) \times p(y=4|-)$.

$= 0.5 \times 0.0463 \times 0.1465 = 0.00339$

$\therefore$ Total unnormalized probablity is =

$(0.0000822 + 0.00339) = 0.0034722$

$\therefore p(+|x=7, y=4) = \frac{0.0000822}{0.0034722}$

$= 0.0237$

$\therefore$ The probability of a customer with $x=7$ and $y=4$ being a positive output is ≈ 0.0237 or 2.37%

9) Calculate the log-likelihood of given dataset.

Data (x): $[2.8, 4.2, 5.3, 2.1]$

Component 1 ($N1$): $\mu_1 = 2.3$, $\sigma_1 = 1.8$

Component 2 ($N2$): $\mu_2 = 6.8$, $\sigma_2 = 2.2$

Prior probabilities $p(N1) = 0.5$, $p(N2) = 0.5$

$$N_1(x) = \frac{1}{\sqrt{2\pi}(1.8)} e^{\frac{-(x-2.3)^2}{2(1.8)^2}}$$

$$N_2(x) = \frac{1}{\sqrt{2\pi}(2.2)} e^{\frac{-(x-6.8)^2}{2(2.2)^2}}$$

The likelihood of a data point is the sum of the weighted densities:

$$L(x) = p(N1) \times N_1(x) + p(N_2) \times N_2(x)$$

$$= 0.5 \times N_1(x) + 0.5 \times N_2(x)$$

data point	N1	N2
2.8	$\frac{0.5 \times 0.217}{0.112} = 0.969$	$\frac{0.5 \times 0.007}{0.112} = 0.031$
4.2	$\frac{0.5 \times 0.124}{0.0825} = 0.752$	$\frac{0.5 \times 0.041}{0.0825} = 0.248$
5.3	$\frac{0.5 \times 0.047}{0.063} = 0.0302$	$\frac{0.5 \times 0.088}{0.0645} = 0.636$
5.5	$\frac{0.5 \times 0.038}{0.063} = 0.302$	$\frac{0.5 \times 0.088}{0.063} = 0.698$
2.1	$\frac{0.5 \times 0.220}{0.1125} = 0.978$	$\frac{0.5 \times 0.005}{0.1125} = 0.022$

New prior probabilities:

$$P(N1)' = \frac{\sum \gamma_{1j}}{n} = \frac{0.969 + 0.752 + 0.364 + 0.302 + 0.978}{5} = 0.678$$

$$P(N2)' = \frac{\sum \gamma_{2j}}{n} = \frac{0.031 + 0.248 + 0.636 + 0.698 + 0.022}{5} = 0.327$$

New means:

$$\mu_1' = \frac{\sum \gamma_{1j} x_j}{\sum \gamma_{1j}} = \frac{(0.969 \times 2.8) + (0.752 \times 4.2) + (0.364 \times 5.3) + (0.302 \times 5.5)}{3.365}$$

$$= \frac{12.355}{3.365} = 3.67$$

$$\mu_2' = \frac{\sum \gamma_{2j} x_j}{\sum \gamma_{2j}} = \frac{(0.031 \times 2.8) + (0.248 \times 4.2) + (0.636 \times 5.3) + (0.698 \times 5.5) + (0.022 \times 2.1)}{1.635}$$

$$= \frac{8.835}{1.635} = 5.40$$

$$\therefore \sigma_1'^2 = \sqrt{\frac{0.969(2.8-3.67)^2 + 0.752(4.2-3.67)^2 + \cdots}{3.365}}$$

$$= \sqrt{0.947} = 0.97$$

$$\sigma_2'^2 = \sqrt{\frac{0.031(2.8-5.40)^2 + 0.248(4.2-5.40)^2 + \cdots}{1.635}}$$

$$= \sqrt{0.297} = 0.54$$

$\therefore$ The Improved components are:

Component 1 (N1):	Component 2 (N2):
$\mu_1 = 3.67,\ \sigma_1 = 0.97$	$\mu_2 = 5.40,\ \sigma_2 = 0.54$
$P(N1) = 0.673$	$P(N2) = 0.327$