Lokesh Patil
20191120041

PAGE No.
DATE / /

Q.1

Sol: Prior probabilities :

P(On time) = 14/20

P(Late) = 2/20

P(Very late) = 3/20

P(cancelled) = 1/20

Posterior probabilities :

i) Attribute - Day

Day	On Time	Late	Very late	cancelled
Weekday	9/14	1/2	3/3	0/1
Saturday	2/14	0/2	0/3	1/1
sunday	1/14	0/2	0/3	0/1
Holiday	2/14	1/2	0/3	0/0

ii) Attribute - Season

Season	On Time	Late	Very Late	Cancelled
Spring	4/14	0/2	0/3	1/1
Summer	8/14	0/2	0/3	0/1
Autumn	2/14	0/2	1/3	0/1
Winter	2/14	2/2	2/3	0/1

iii) Attribute - Fog

Fog	On Time	Late	Very Late	Cancelled
None	5/14	0/2	0/3	0/1
High	4/14	1/2	1/3	1/1
Normal	5/14	1/2	2/3	0/1

iv) Attribute – Rain

Rain	On Time	Late	Very Late	Cancelled
None	6/14	1/2	1/3	0/1
Slight	6/14	1/2	0/3	0/1
Heavy	2/14	0/2	2/3	1/1

Using Naive Bayes formulae –

$P_{NB}(\text{On Time}) = P(\text{On Time}) \times P(\text{Weekday} / \text{On Time}) \times P(\text{Winter} / \text{On Time}) \times P(\text{High} / \text{On Time}) \times P(\text{None} / \text{On Time})$

$= \frac{14}{20} \times \frac{9}{14} \times \frac{2}{14} \times \frac{4}{14} \times \frac{6}{14}$

$= 0.0079$

$P_{NB}(\text{Late}) = P(\text{Late}) \times P(\text{Weekday} / \text{Late}) \times P(\text{Winter} / \text{Late}) \times P(\text{High} / \text{Late}) \times P(\text{None} / \text{Late})$

$= \frac{2}{20} \times \frac{1}{2} \times \frac{2}{2} \times \frac{1}{2} \times \frac{1}{2}$

$= 0.0125$

$P_{NB}(\text{Very Late}) = P(\text{Very Late}) \times P(\text{Weekday} / \text{Very Late}) \times P(\text{Winter} / \text{Very Late}) \times P(\text{High} / \text{Very Late}) \times P(\text{None} / \text{Very Late})$

$= \frac{3}{20} \times \frac{3}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3}$

$= 0.011$

$P(\text{Cancelled}) = \frac{1}{20} \times \frac{0}{1} \times \frac{0}{1} \times \frac{1}{1} \times \frac{0}{1}$

$= 0$

Here P(Late) is highest
∴ The correct classification is Late

Q. 2

Sol. We have to test hypothesis that gender and preferred reading are independent. that means there is no correlation between them.

Using Chi-square test
Contingency table -

	Male	Female
Fiction	250 (90)	200 (360)
Non-fiction	50 (210)	1000 (840)

Degrees of Freedom = (2 - 1) × (2 - 1)
= 1

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$$

$$= \frac{(250 - 90)^2}{90} + \frac{(50 - 210)^2}{210} + \frac{(200 - 360)^2}{360} + \frac{(1000 - 840)^2}{840}$$

$$= 507.94$$

For degree of freedom 1, significance - 0.01
χ^2 value needed to reject hypothesis is 6.635.
The received value is above this value. Thus, we reject the hypothesis that the gender and preferred reading are independent.