Assignment 1 D.A. BE ETRX

Eeshaan Asodekar 2019110002

(Q1) Finding prior and posterior probabilities

prior prorities :
P(on time) = 14/20
P (late) = 2/20
P (v late) = 3/20
P (cancelled) = 1/20

Computing posterior probabilities:

→ for 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	0/1

→ For 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	0/1
normal	5/14	1/2	2/3	0/1

→ for "season"

Season	On time	late	very late	Cancelled
Spring	4/14	0/2	0/3	1/1
Summer	6/14	0/2	0/3	0/1
Autumn	2/14	0/2	1/3	0/1
Winter	2/14	2/2	2/3	0/1

→ for "Rain"

Rain	Ontime	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

Computing the probabilities for the outcome ~~for~~ the <weekday, winter, high, none>

P_{NB}(on time)

$= P(\text{ontime}) \times P(\text{weekday}|\text{ontime}) \times P(\text{winter}|\text{ontime}) \times P(\text{high}|\text{ontime}) \times P(\text{None}|\text{ontime})$

$= \frac{14}{20} \times \frac{9}{14} \times \frac{2}{14} \times \frac{4}{14} \times \frac{6}{14} = 0.0079$

P_{NB}(late)

$= P(l)\ P(wd|l)\ P(w|l)\ P(h|l)\ P(\text{none}|l)$

$= \frac{2}{20} \times \frac{1}{2} \times \frac{2}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 0.0125$

P_{NB}(very late)

$= P(vl)\ P(wd|vl)\ P(w|vl)\ P(h|vl)\ P(\text{none}|vl)$

$= \frac{3}{20} \times \frac{3}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} = 0.0111$

P_{NB}(cancelled)

$= P(c)\ P(wd|c)\ P(w|c)\ P(h|c)\ P(\text{none}|c)$

$= \frac{1}{20} \times \frac{0}{1} \times \frac{0}{1} \times \frac{1}{1} \times \frac{0}{1} = 0$

P_{NB}(late) is highest

P_{NB}(late) > { P(V.l) > P(ot) > P(c) }

∴ given i/p tuple is classified as "Late"

(Q2) total = 1500 people

To test hypothesis: gender and reading preferences are independent, and uncorrelated

Using X^2 correlation test with contingency table (2×2) ~~and~~ × (2-1) degrees of freedom

$$X^2 = \sum_{i}^{2} \sum_{j}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where O_{ij} → observed freq.
E_{ij} → expected freq.

X^2 =

$$\frac{(250-90)^2}{90} + \frac{(50-210)^2}{210} + \frac{(200-360)^2}{360} + \frac{(1000-840)^2}{840} = 507.9365$$

At 0.01 α, and df = 1, X^2 rejection boundary value is 10.828

Page No.

Date

Since $507.94 > 10.828$; we can reject the hypothesis that gender and preferred reading are independent and therefore, we conclude that there is some correlation between them.