DA Assignment-1

Name:- Aditi Chaturvedi
UID: 2019120011

Ans 2:) ~~Proof~~

H_0: Preferred reading and gender are not correlated in a group.

H_1: Preferred reading and gender are correlated in a group.

→ Given frequencies

$e11 = [\text{count(male)} \times \text{count(fiction)}]/N$

$\therefore e11 = 90$

$e12 = 360$

$e21 = 210$

$e22 = 840$

→ Compute χ^2

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})}{(\text{Expected})}$$

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

$$\chi^2 = 284.44 + 121.90 + 71.11 + 30.48$$

$$= 507.93$$

→ For 2×2 table
degree of freedom = $(2-1)(2-1) = 1$

→ For 1 degree of freedom, χ^2 value needed to reject the hypothesis at 0.001 significant level is 10.828 (referred from χ^2 distribution table)

→ We observe that the computed value is greater therefore we reject the null hypothesis i.e
Preferred reading and gender are not correlated in a group.

→ Conclusion: Preferred reading and gender are correlated in the group.

Naive Bayes Classification

Ans1) P(Class = On Time) = 14/20
P(Class = Late) = 2/20
P(Class = Very Late) = 3/20
P(Class = Cancelled) = 1/20

(I) Days

Days	On Time	Late	Very Late	Cancelled
Weekday	9/14	1/2	3/3	0/1
Holiday	2/14	1/2	0/3	0/1
Saturday	2/14	0/2	0/3	1/1
Sunday	1/14	0/2	0/3	0/1

(II) 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
Winter	2/14	2/2	2/3	0/1
Autumn	2/14	0/2	1/3	0/1

(III)

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

(IV)

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

→ < Day = WeekDay, Season = Winter, Fog = High, Rain = None >

→ $V_{NB} = \underset{v_j \in \{yes, no\}}{\text{argmax}} \; P(v_j) \cdot \prod_i (a_i | v_j)$

→ $V_{NB} = \underset{v_j \in \{\text{On Time, Late, Very Late, Cancelled}\}}{\text{argmax}} \; P(v_j) \prod_i (a_i | v_j)$

→ $V_{NB}(\text{On Time}) = P(\text{On Time}) \cdot P(\text{Weekday} | \text{On Time}) \cdot P(\text{Winter} | \text{On Time}) \cdot P(\text{High} | \text{On Time}) \cdot P(\text{None} | \text{On Time})$

→ $V_{NB}(\text{On Time}) = \frac{14}{20} \times \frac{9}{14} \times \frac{2}{14} \times \frac{4}{14} \times \frac{6}{14}$

$= 0.0078717$

$= 7.8717 \times 10^{-3}$ //

~~→ VNB (Late)~~

→ $V_{NB}(\text{Late}) = P(\text{Late}) \cdot P(\text{Weekday} | \text{Late}) \cdot P(\text{Winter} | \text{Late}) \cdot P(\text{High} | \text{Late}) \cdot P(\text{None} | \text{Late})$

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

$= 0.0125$ //

→ $V_{NB}(\text{Very Late}) = P(\text{Very Late}) \cdot P(\text{Weekday} | \text{Very Late}) \cdot P(\text{Winter} | \text{Very Late}) \cdot P(\text{High} | \text{Very Late}) \cdot P(\text{None} | \text{Very Late})$

$= \frac{3}{20} \times \frac{3}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} = 0.01111$

Date:

→ VNB(Cancelled) = P(Cancelled) · P(WeekDay | Cancelled) · P(Winter | Cancelled) · P(High | Cancelled) · P(None | Cancelled)

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

VNB(On Time) = 0.25004

VNB(Late) = 0.397056

VNB(Very Late) = 0.352585

VNB(Cancelled) = 0.

→ Since VNB is the greatest
∴ The instance will be categorised under Late.