DA Assignment (Rushil Patel) (201914004-7)

Q1] Calculating all prior and posterior probabilities

~~Prior probabilities:~~

| Attribute		Ontime	Late	Very late	Cancelled
Day:	Weekday	9/14 = 0.64	1/2 = 0.5	3/3 = 1	0/1 = 0
	Saturday	2/14 = 0.14	1/2 = 0.5	0/3 = 0	1/1 = 1
	Sunday	1/14 = 0.07	0/2 = 0	0/3 = 0	0/1 = 0
	Holiday	2/14 = 0.14	0/2 = 0	0/3 = 0	0/1 = 0
Season:	Spring	4/14 = 0.29	0/2 = 0	0/3 = 0	0/1 = 0
	Summer	6/14 = 0.43	0/2 = 0	0/3 = 0	0/1 = 0
	Autumn	2/14 = 0.14	0/2 = 0	1/3 = 0.33	0/1 = 0
	Winter	2/14 = 0.14	2/2 = 1	2/3 = 0.66	0/1 = 0
Fog:	None	5/14 = 0.36	0/2 = 0	0/3 = 0	0/1 = 0
	High	4/14 = 0.29	1/2 = 0.5	1/3 = 0.33	1/1 = 1
	Normal	5/14 = 0.36	1/2 = 0.5	2/3 = 0.67	0/1 = 0
Rain:	None	5/14 = 0.36	1/2 = 0.5	1/3 = 0.33	0/1 = 0
	Slight	8/14 = 0.57	0/2 = 0	0/3 = 0	0/1 = 0
	Heavy	1/14 = 0.07	1/2 = 0.5	2/3 = 0.67	1/1 = 1

Prior probability	14/20 = 0.70	2/20 = 0.10	3/20 = 0.15	1/20 = 0.05
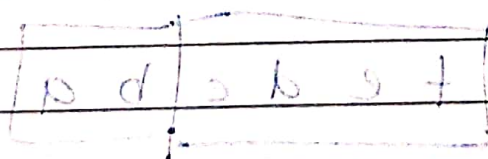
Instance : Weekday, Winter, High, None ??

Case 1 Class = on time

$$= (0.7)(0.64)(0.14)(0.29)(0.36)$$
$$= 6.547 \times 10^{-3}$$

Case 2 Class = Late

$$= (0.1)(0.50)(1.0)(0.5)(0.5)$$
$$= 0.0125$$

or

Case 3 Class = (Very late)

$$= (0.15)(1.0)(0.67)(0.33)(0.33)$$
$$= 0.0109$$

Case 4 Class = Cancelled

$$= (0.05)(0.0)(0.0)(1.0)(0)$$
$$= 0$$

Case 2 is strongest and Hence instance will be categorized under class "Late".

Q2] Ho: Preffered reading and gender are not correlated correlated in the grp.

H₁: Both are correlated

⇒ Computing X^2 value,

$$X^2 = \frac{\text{count}(A = a_i) \times \text{count}(B = b_j)}{n}$$

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

$$= 284.44 + 121.9 + 71.11 + 30.48$$

$$= 507.93$$

For (2×2) table degree of freedom are

$(2-1)(2-1) = 1$

For 1 degree of freedom, X^2 value needed to reject the hypothesis at 0.001 significance level is 10.828 (took from X^2 distribution)

Since, the computed value is above this, we can reject the null hypothesis that gender and preferred reading are independent.

∴ We conclude that 2 attributes are correlated for given group

formula: $$X^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$$