EDA Assignment

Q.2 In this prob, we will test the hypothesis that gender and prefered reading have no corr b/w them & they are independent.

Sol By using $\chi^2$-test.

DOF = (2-1) × (2-1)

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{[a_{ij} - e_{ij}]^2}{e_{ij}} \Rightarrow \chi^2 = \frac{[250-90]^2}{90} + \frac{[50-210]^2}{210} + \frac{[200-360]^2}{360} + \frac{[1000-840]^2}{840}$$

$$= 507.93$$

$507.93 > 2.706$ [$\chi^2$ value for dof = 1 & $\alpha$ or $\alpha = 0.1$]

∴ The gender and prefered reading are strongly correlated.

So we reject the null hypothesis of independence at a confidence level of 0.1

Q.1 Naive Bayesian Classifier

| | Attribute | On Time | Late | Very Late | Can |
|---|---|---|---|---|---|
| Day | Week day | 0.64 | 0.5 | 1 | 0 |
| | Sat | 0.14 | 0.5 | 0 | 1 |
| | Sun | 0.07 | 0 | 0 | 0 |
| | Holiday | 0.14 | 0 | 0 | 0 |
| Season | Spring | 0.29 | 0 | 0 | 0 |
| | Summ | 0.43 | 0 | 0 | 0 |
| | Aut | 0.14 | 0 | 0.37 | 0 |
| | wint | 0.14 | 1 | 0.67 | 0 |
| Fog | None | 0.36 | 0 | 0 | 0 |
| | High | 0.29 | 0.5 | 0.33 | 1 |
| | Normal | 0.36 | 0.5 | 0.67 | 0 |
| Rain | None | 0.36 | 0.5 | 0.33 | 0 |
| | Slight | 0.57 | 0 | 0 | 0 |
| | Heavy | 0.37 | 0.5 | 0.67 | 1 |
| Prior Prob | | 0.70 | 0.10 | 0.15 | 0.05 |

For the instance in the question
<weekday, winter, high, None>

$P_{NB}(\text{on time}) = \frac{14}{20} \times \frac{9}{14} \times \frac{2}{14} \times \frac{4}{14} \times \frac{5}{9} = 0.0079$

Similarly,

$P_{NB}(\text{late}) = 0.0125$
$P_{NB}(\text{very late}) = 0.0111$
$P_{NB}(\text{Cancelled}) = 0$

$P_{NB}(\text{late})$ has the highest value. ∴ the instance belongs to late class