DA Assignment 1

Anuj Chavan
201912 0012

Q.1 Calculating prior probabilities

$P(\text{on-time}) = \frac{14}{20}$ $P(\text{late}) = \frac{2}{20}$ $P(\text{Cancelled}) = \frac{1}{20}$

$P(\text{very late}) = \frac{3}{20}$

Day	On time	late	Very-late	Cancelled
Weekday	9/14	1/2	3/3	0
Holiday	2/14	1/2	0	0
Saturday	2/14	0	0	1
Sunday	1/14	0	0	0

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

Fog	On-time	late	Very-late	Cancelled
High	4/14	1/2	1/3	1
Normal	5/14	1/2	2/3	0
None	5/14	0	0	0

Rain	On time	Late	Very-late	Cancelled
None	6/14	1/2	1/3	0
Slight	6/14	1/2	~~2/3~~ 0	0
Heavy	2/14	0	0 2/3	1

Instance = {Day = weekday, season = winter, Fog = high, Rain = None}

$$V_{NB} = \underset{v_j = \{\text{ontime, late, very-late, cancelled}\}}{\operatorname{argmax}} P(v_j) \prod_i P(a_i \mid v_j)$$

$$= \underset{v_j = \{\text{ontime, late, very-late, cancelled}\}}{\operatorname{argmax}} P(v_j) \prod P(\overset{\text{weekday}}{\cancel{\text{Day}}} \mid v_j)\, P(\overset{\text{winter}}{\cancel{\text{Season}}} \mid v_j)\, P(\overset{\text{high}}{\cancel{\text{Fog}}} \mid v_j)\, P(\overset{\text{None}}{\cancel{\text{Rain}}} \mid v_j)$$

$$V_{NB}(\text{on-time}) = P(\text{ontime}) \times P(\text{Weekday} \mid \text{ontime}) \times P(\overset{\text{winter}}{\cancel{\text{Season}}} \mid \text{Ontime}) \times P(\overset{\text{High}}{\cancel{\text{Fog}}} \mid \text{Ontime}) \times P(\overset{\text{None}}{\cancel{\text{Rain}}} \mid \text{Ontime})$$

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

$$\boxed{V_{NB}(\text{ontime}) \simeq 0.0078}$$

$$V_{NB}(\text{late}) = \frac{2}{20} \times P(\text{late}) \times P(\text{weekday} \mid \text{late}) \times P(\text{winter} \mid \text{late}) \times P(\text{high} \mid \text{late}) \times P(\text{None} \mid \text{late})$$

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

$$\boxed{V_{NB}(\text{late}) = 0.0125}$$

$$V_{NB}(\text{very-late}) = P(\text{very late}) \times P(\text{weekday} \mid \text{very-late}) \times P(\text{winter} \mid \text{very-late}) \times P(\text{High} \mid \text{very-late}) \times P(\text{None} \mid \text{very-late})$$

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

$$\boxed{V_{NB}(\text{very-late}) = 0.0111}$$

$$V_{NB}(\text{cancelled}) = P(\text{cancelled}) \times P(\cancel{\text{weekday}} \mid \text{cancelled}) \times P(\text{Winter} \mid \text{cancelled}) \times P(\text{High} \mid \text{cancelled}) \times P(\text{None} \mid \text{cancelled})$$

$$\cancel{V_{NB}(\text{ca}} = \frac{1}{20} \times 0 \times 0 \times 1 \times 0 = 0$$

$$\boxed{V_{NB}(\text{cancelled}) = 0}$$

Therefore the given instance can be classified into the class of <u>'late'</u>.

Q.2

H_0 : The Preffered reading and gender are not correlated

H_1 : Both are correlated.

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

$$e_{ij} = \frac{\text{count}(A = a_i) \times \text{count}(B = b_j)}{N}$$

$$\chi^2 = \frac{(250-90)^2}{90} + \frac{(200-360)^2}{360} + \frac{(50-210)^2}{210} + \frac{(1000-840)^2}{840} = \underline{507.93}$$

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

Looking at the χ^2 table,
the value needed to reject hypothesis at 0.001 significance level is 10.828.

Since the value obtained is above this value,
we reject the hypothesis.

$\therefore$ We conclude that both the attributes are correlated