Jayesh Vikas Bane 2019120006

DA Assigment

Problem - 1.

We have 4 attributes => Day, Season, Fog, Rain.

We have 4 class categories => OnTime, Late, Very late, Cancelled.

Applying Naive Bayes Classifier.

Prior Probabilities for class categaries.

$P(\text{On Time}) = 14/20$
$P(\text{Late}) = 2/20$
$P(\text{Very Late}) = 3/20$
$P(\text{Cancelled}) = 1/20$

Posterior Probabilites.

For attribute 'Day':

Day \ Class	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 'Season':

Season \ Class	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 attribute 'Fog':

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

For attribute 'Rain':

Rain \ Class	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

Unseen Instance = <Weekday, Winter, High, None>

Case 1 => On Time

$= \cancel{\frac{14}{20}} \times$

$= P(\text{On Time}) \times P(\text{Weekday} / \text{On time}) \times P(\text{Winter} | \text{On time})$
$\times P(\text{High} | \text{On time}) \times P(\text{None} | \text{On time})$

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

$= 7.87 \times 10^{-3} = 0.00787$

Similarly

Case 2 => Late.

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

Case 3 => Very Late.

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

4

Case 4 => Cancelled.

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

Probability for Late is the highest and & hence the unseen instance is classified as late. Ans.

Problem 2

Attributes => Gender and Preferred Reading

	Male	Female	Total
Fiction	250 (90)	200 (360)	450
Non-fiction	50 (210)	1000 (840)	1050
Total	300	1200	1500

Using χ^2 Test on given data

$$\chi^2 = \sum \left[\frac{(\text{observed} - \text{expected})^2}{\text{expected}} \right]$$

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

$$\chi^2 = 507.93$$

Degree of freedom for a table of $m \times n$ is $(m-1) \times (n-1)$

$\therefore$ DF $= (2-1) \times (2-1) = 1$.

Assuming significance level $(\alpha) = 1\%$.

From chi-square table for DF $= 1$ & $\alpha = 1\%$. the value is 6.635.

Since $507.93 > 6.635$, we reject the hypothesis that gender and preferred reading are independent at $\alpha = 1\%$.