PAGE No.
DATE / /

DA Assignment 1

Que. 1]. We have to apply Naive Bayte's Classifier to map input tuples into accurate to classes.

We know that table class has 4 diff. values, we need to find all the prior & posterios probabilities.

$P(\text{On Time}) = \frac{14}{20}$, $P(\text{Late}) = \frac{2}{20}$

$P(\text{Very Late}) = \frac{3}{20}$, $P(\text{Cancelled}) = \frac{1}{20}$

We need to calculate posterior probalitities.

For attribute → 'Day'

$P(\text{Weekday} / \text{On Time}) = \frac{9}{14}$

$P(\text{Weekday} / \text{Late}) = \frac{1}{2}$

$P(\text{Weekday} / \text{Very Late}) = \frac{3}{13}$

$P(\text{Weekday} / \text{Cancelled}) = \frac{0}{1}$

In similar way we can calculate posterior probabilities for all other value of attribute.

Day	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	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	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	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

By applying Naive Bayes formula,

$$P_{NB}(\text{On Time}) = P(\text{On Time}) \times P(\text{Weeekday} / \text{On Time}) \times P(\text{Winter} / \text{On Time}) \times P(\text{High} / \text{On Time}) \times P(\text{None} / \text{On Time})$$

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

$$= 0.0079$$

Applying Naive Bayes formula for 'Late', 'Very Late' & 'Cancelled'

$$P_{NB}(\text{Late}) = \frac{2}{20} \times \frac{1}{2} \times \frac{2}{2} \times \frac{1}{2} \times \frac{1}{2}$$

$$= 0.0125$$

$$P_{NB}(\text{Cancelled}) = \cancel{3}\ \frac{1}{20} \times \frac{0}{1} \times \frac{0}{1} \times \frac{0}{1} \times \frac{1}{1}$$

$$= 0$$

$$P_{NB}(\text{Very late}) = \frac{3}{20} \times \frac{3}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3}$$

$$= 0.011$$

As P_{NB} (Late) is highest among all, therefore, Late is the correct classification.

By using this method, we can ~~d~~ also find any other unseen instances predictions.

PAGE No.	
DATE	/ /

Que. 2] → We have to test hypothesis that gender preferred reading are independent. i.e. there is no correlation between them.

By using χ^2 test

The size of countinger error table is 2×2.

	Male	Female
Fiction	250 (90)	200 (360)
Non-fiction	50 (210)	1000 (840)

Degree of freedom
$\Rightarrow (2-1)\times(2-1)$
$= 1$

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

where,
$e_{ij} \rightarrow$ Expected frequency
$o_{ij} \rightarrow$ Observed frequency

$$\chi^2 = \frac{(250-70)^2}{90} + \frac{(50-210)^2}{210} + \frac{(200-360)^2}{360} + \frac{(1000-840)^2}{840}$$

$= 507.937$

PAGE No.
DATE / /

By referring the table, for degree of freedom 1 significant 0.01.
χ^2 value needed to reject hypothesis is 6.675

As the value we received is above this value, so we can reject the hypothesis that gender & preferred reading are independent and we can conclude that two attributes are correlated.