Assignment - I

Q.1.]

AIR TRAFFIC DATA

Swapnil Mandewalli
2019120037
BE - EXTC

ATTRIBUTES :	DAY
	SEASON
	FOG
	RAIN

CATEGORIES :	ON TIME	14/20
	LATE	2/20
	VERY LATE	3/20
	CANCELLED	1/20

NAIVE BAYESAIN CLASSIFIER

$$P(y|x_1 \ldots x_n) = \frac{P(x_1|y)\,P(x_2|y) \ldots P(x_n|y)\,P(y)}{P(x_1)\,P(x_2)\,P(x_3) \ldots P(x_n)}$$

TABULATION OF ALL PROBABILITIES

		CLASS			
	ATTRIBUTE	ON TIME	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 = 0.67	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
		14/20 = 0.7	2/20 = 0.1	3/20 = 0.15	1/20 = 0.05

Swapnil Mandavilli

INSTANCE

WEEKDAY WINTER HIGH NONE

CASE 1: Class: ON TIME $\Rightarrow 0.7 \times 0.64 \times 0.14 \times 0.29 \times 0.36 = 0.0065$

CASE 2: Class: LATE $\Rightarrow 0.1 \times 0.5 \times 1 \times 0.5 \times 0.5 = 0.0125$

CASE 3: Class: Very Late $\Rightarrow 0.15 \times 1 \times 0.67 \times 0.33 \times 0.33 = 0.0111$

CASE 4: Class: Cancelled $\Rightarrow 0.05 \times 0 \times 0 \times 1 \times 0 = 0$

Case 2 that is class: LATE is the strongest, since 0.0125 is greater than all the other 3 values.

∴ The correct classification for the given instance is late.

ANS: LATE

Swapnil Mandavilli
2019120037
BE - EXTC

Q.2.]

- Null Hypothesis (H_0): Preferred reading and gender are not correlated in the group.
- Alternate Hypothesis (H_a): Preferred reading and gender are correlated in the group.
- Computing the χ^2 value,

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

As per the table,

$O_{11} = 250$	$e_{11} = 90$
$O_{12} = 200$	$e_{12} = 360$
$O_{21} = 50$	$e_{21} = 210$
$O_{22} = 1000$	$e_{22} = 840$

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

$$= 284.44 + 121.90 + 71.11 + 30.48 = 507.93.$$

- For 2x2 table, degree of freedom are $(2-1)(2-1) = 1$
- Degree of freedom value: 1 } 10.828 (from χ^2 distribution table)
  Significance level: 0.001 }
- Since, the computed value is greater than permitted value ($507.93 > 10.828$), we reject the null hypothesis.

∴ We conclude that preferred reading and gender are correlated.