Q1] There are 4 types of attributes - Day, Season, Fog, Rain with 20 tuples.

Also, there are 4 categories - on time, Late, Very late, Cancelled.

Naive Bayesian Classifier can be used to map any unseen tuple into an accurate class.

Prior Probabilities :-

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

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

Posterior Probabilities

For attribute "Day" :-

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

for the given instance.
[weekday, Winter, High, None]

P(ON time) = P(weekday | ON time) × P(winter | ON time) × P(High | ON time) × P(None | ON time) × P(ON time)

$$= \frac{9}{14} \times \frac{2}{14} \times \frac{4}{14} \times \frac{6}{14} \times \frac{14}{20} = 0.0079$$

P(Late) = P(weekday | Late) × P(winter | late) × P(High | Late) × P(None | Late) × P(Late)

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

P(very late) = P(weekday | very late) × P(winter | very late) × P(High | very late) × P(None | very late) × P(very late)

$$= \frac{3}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{3}{20} = 0.0111$$

P(cancelled) = P(weekday | cancelled) × P(winter | cancelled) × P(High | cancelled) × P(None | cancelled) × P(cancelled)

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

Here P(Late) is highest.

∴ The correct classification is Late.

Any other unseen instances' ~~can be found~~ prediction can be found out by this method.

Q2] To test hypothesis ~~between~~ that gender & preferred reading are independent. This means there is no correlation between them.

We can use Chi-Square test.

The contingency table size is 2×2 which is given to us.

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

Degree of freedom = (2-1)×(2-1) = 1.

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

O_{ij} → observed Frequency

e_{ij} → Expected Frequency.

$$\chi^2 = \frac{(250-90)^2}{90}$$

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

$$= 507.94$$

Referring the table, for Degree of freedom 1 & Significance 0.01, χ^2 value needed to reject hypothesis is 6.635

Our received value is well above this value. Therefore we can reject the hypothesis that gender and preferred reading are independant and we can conclude that the two attributes are correlated.