MAX. MARKS: 25

COURSE NAME: Data Warehousing and Data Mining
COURSE INSTRUCTORS: Dr. Pardeep Kumar   MAX. TIME: 1 Hour 30 Minutes

Note: (a)All questions are compulsory.
(b)Marks are indicated against each question in square brackets.
(c) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

1. Suppose we are interested in analyzing transactions at AllElectronics with respect to the purchase of computer games and videos. Let game refer to the transactions containing computer games, and video refer to those containing videos. Of the 10,000 transactions analyzed, the data show that 6000 of the customer transactions included computer games, while 7500 included videos, and 4000 included both computer games and videos.
   (i)   Make the contingency table of your analysis
   (ii)   Compute the degree of freedom.
   (iii)   Illustrate whether the hypothesis "game and video attributes are independent" is true or false for level of significance ($\alpha=0.001$). Note that the chi square tabular value for $\alpha=0.001$ and your degree of freedom is 10.28.

[2+1+5] [CO-3]

2. Consider the weather data set given below:

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Make a classification tree using ID-3 approach to predict whether the game will be played or not on a particular day. Also compute the accuracy and coverage of your inference drawn for each rule.

[8+2] [CO-4]

Q3. Explain IQR method to handle noise/outliers. Find the noise/outliers in the following attribute using IQR method: Age (10, 16, 11, 7, 4, 4, 4, 12, 17, 25, 14, 0).   [CO-2] [3+4]