1. (a) Define Predictive Analytics and explain its relationship with Data Mining. **2**

(b) For the following meetings, identify the CRISP-DM phase.

(i) The data mining project manager meets with the data warehousing manager to discuss how the data will be collected. **2**

(ii) The data mining project manager meets with the production line supervisor, to discuss implementation of changes and improvements.

(iii) Managers want to know by next week whether deployment will take place. Therefore, analysts meet to discuss how useful and accurate their model is.

(iv) The analysts meet to discuss whether the neural network or decision tree models should be applied.

(c) Differentiate between classification and prediction. **2**

2. (a) The following are the attributes of a university student. State if the data are nominal, ordinal, interval or ratio scaled. **2**

(i) Gender – Male

(ii) Class – Third Year

(iii) Blood Type – A+

(iv) Body Temperature – 36° C

(b) Find Q1, Q2, Q3, and IQR for the following data set. **2**

{1,6,7,8,8,11,12,13,14,16,22,35}

(c) Draw a box-and-whisker plot for the data given in Q. 2 (b). **2**

3. (a) Compute skewness for the given data and comment on it. **2**

{5, 204, 210, 8, 10, 11, 55, 65, 72, 92, 35, 50, 108, 150, 187, 13, 15, 215}

(b) Find min-max normalized stock price (new range: [0–1]) for following stock data:

Stock Price (₹): 20, 52, 189, 421, 32 **2**

(c) Use smoothing by bin means and boundaries to smooth the following data using a bin depth of 6. **2**

{11, 13, 13, 15, 15, 16, 19, 20, 20, 20, 21, 21, 22, 23, 24, 30,

40, 45, 45, 45, 71, 72, 73, 90}

4. (a) Compare and contrast Hypothesis Testing and Exploratory Data Analysis (EDA). **2**

(b) Consider a sample of N=200 beer-drinkers. For each drinker we have information on sex (variable X, taking on 2 possible values: "Male" and "Female") and preferred category of beer (variable Y, taking on 3 possible values: "Light", "Regular", "Dark"). A contingency table for these data is given below: **2**

	Light	Regular	Dark	Total
Male	20	40	50	110
Female	50	20	20	90
Total	70	60	70	200

Illustrate above data with the help of clustered bar chart.

(c) Discuss the need for dimensionality reduction in Data Mining. **2**

5. (a) Consider following sample data: **2**

Observation	Feature (X₁)	Feature (X₂)
A	4	11
B	8	4
C	13	5
D	7	14

Compute the standardize matrix using Z-score standardization.

(b) Compute covariance matrix for the data obtained in Q. 5(a) after standardization. **2**

(c) Compute the eigen values and eigen vectors for Q. 5(a). **2**

End of Questions