Hall Ticket No | | | | | | | | | | | Question Paper Code: ACDC13

INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)
Dundigal-500043, Hyderabad

B.Tech VII SEMESTER END EXAMINATIONS (REGULAR) - DECEMBER 2023
Regulation: UG-20
DATA MINING

Time: 3 Hours **CSE (DATA SCIENCE)** **Max Marks: 70**

Answer ALL questions in Module I and II
Answer ONE out of two questions in Modules III, IV and V
All Questions Carry Equal Marks
All parts of the question must be answered in one place only

MODULE – I

1. (a) Describe different types of data and their significance in data mining. Compare data mining and data warehouse. [BL: Understand| CO: 1|Marks: 7]

   (b) Consider the following data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

   i) Use min-max normalization to transform the value 35 for age on to the range [0.0, 1.0].
   ii) Use z-score normalization to transform the value 35 for age, where the standard deviation of age is 12.94 years.
   iii) Use normalization by decimal scaling to transform the value 35 for age.
   iv) Comment on which method you would prefer to use for the given data, giving reasons as to why. [BL: Apply| CO: 1|Marks: 7]

MODULE – II

2. (a) Identify the approaches for mining multi level association rules from the transactional databases. Give relevant examples. [BL: Understand| CO: 2|Marks: 7]

   (b) Apply apriori algorithm to find frequent itemsets from the following transactional database given in Table 1. Let min_sup = 30%. [BL: Apply| CO: 2|Marks: 7]

Table 1

TID	Items_bought
1	Pen, notebook, ruler
2	Pencil, eraser, sharpener
3	Pen, ruler, chart, sharpener
4	Pencil, clip, eraser
5	Ruler, pin, story book, pen
6	Marker, chart, sketchpens

MODULE – III

3. (a) Outline the major issues regarding classifications and predictions. Differentiate between classification and prediction in data mining. [BL: Understand| CO: 3|Marks: 7]

   (b) How neural networks can be used for data classification? Which algorithm is suitable? Explain them with example? [BL: Apply| CO: 3|Marks: 7]

4. (a) State classification problem and discuss a general approach to solve classification problem. [BL: Understand| CO: 4|Marks: 7]

   (b) Explain decision tree induction algorithm for classifying data tuples and with suitable example. [BL: Understand| CO: 4|Marks: 7]

MODULE – IV

5. (a) Elucidate density-based and grid-based methods in clustering. Compare partitioning methods with hierarchical methods in clustering. [BL: Understand| CO: 5|Marks: 7]

   (b) Suppose that the data mining task is to cluster the following eight students given in Table 2 into three clusters, the distance function is Manhattan.Assign record 1,2,3 as the centroid of each cluster respectively.Use the k-means algorithm to show the final three clusters. [BL: Apply| CO: 5|Marks: 7]

Table 2

RecordID	Height(cms)	Weight(kgs)
1	145	35
2	165	55
3	170	90
4	135	60
5	140	50
6	160	75
7	150	40
8	155	65

6. (a) List the evaluation metrics used for clusters. Analyze the effectiveness of grid-based methods in clustering. [BL: Understand| CO: 5|Marks: 7]

   (b) Appraise the importance of outlier detection and its application. Explain any one approach for outlier detection. [BL: Apply| CO: 5|Marks: 7]

MODULE – V

7. (a) Summarize about hierarchy of categories in text mining. Compare text mining with web content mining using lucid examples. [BL: Understand| CO: 6|Marks: 7]

   (b) List the applications of web usage mining. Give a brief note on PageRank algorithm used in web structure mining. [BL: Apply| CO: 6|Marks: 7]

8. (a) Discuss various kinds of patterns to be mined from web/server logs in web usage mining. [BL: Understand| CO: 6|Marks: 7]

   (b) How to convert unstructured text in to features in text mining? Discuss the data mining tasks applicable to text databases. [BL: Apply| CO: 6|Marks: 7]

– ○ ○ ◯ ○ ○ –