



U.S.N

--	--	--	--	--	--	--	--	--	--

P.E.S. College of Engineering, Mandya - 571 401

(An Autonomous Institution affiliated to VTU, Belagavi)

Sixth Semester, B.E. - Computer Science and Engineering

Semester End Examination; May / June - 2018

Data Warehousing and Data Mining

Time: 3 hrs

Max. Marks: 100

Note: Answer FIVE full questions, selecting ONE full question from each unit.

UNIT - I

- | | | |
|------|--|----|
| 1 a. | Compare and contrast OLTP and data warehouse. | 5 |
| | b. What is data warehouse? What are the goals of a data warehouse? | 5 |
| | c. Describe the data cube technology with an example. | 10 |
| 2 a. | With a neat block diagram, explain the working principle of OLTP and OLAP architectures. | 10 |
| | b. i) Differentiate between data warehouse and operational database | |
| | ii) What are the characteristics of the OLTP and the basic data warehouse environments as they relate to information delivery needs? | 10 |

UNIT - II

- | | | |
|------|--|----|
| 3 a. | Briefly explain the data mining strategies. | 10 |
| | b. Discuss the various types of data attributes, with suitable examples. | 10 |
| 4 a. | Explain the importance of data pre-processing. Further, explain some of the data pre-processing technologies used to handle text data. | 10 |
| | b. With a proper justification, explain the importance of dimensionality reduction techniques (DRT). With a suitable example, explain any one DRT. | 10 |

UNIT - III

- | | | |
|------|--|----|
| 5 a. | What is classification? List and explain the important stages related to classify the video data. | 10 |
| | b. Differentiate between rule based ordering scheme and class based ordering scheme. What are the steps to be considered to improve the classification accuracy? | 10 |
| 6 a. | With a neat mathematical formulations, explain the working principle of Naïve Bayes classifier | 10 |
| | b. What is over fitting and under fitting? With an example, explain the working principle of confusion matrix. | 10 |

UNIT - IV

- | | | |
|------|--|----|
| 7 a. | Explain the various alternative methods for generating frequent item sets. | 10 |
| | b. Explain the working principle of Apriori algorithm with an example. | 10 |

Contd...2

- 8 a. Illustrate the limitations of support-confidence frame work for evaluation of an Association rule. 10
- b. Define support and confidence. Suppose the support of milk is 70%, Support for sugar is 10% and support for caviar is 0.04%. Given $h_c = 0.01$, Is the frequent item set {milk, sugar, caviar} the cross support pattern? Justify. 10

UNIT - V

- 9 a. What is unsupervised learning? List the various characteristics that data should exhibit to achieve good clusters. 10
- b. Explain the importance of cluster validation. Explain any one internal and external cluster validation technique. 10
- 10 a. Explain the working principle of hierarchical clustering methods. 10
- b. Consider a set of points, explain the working principle k-means clustering for the same set of points. Debate on its complexity issues. 10

* * * *