MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL

Paper Code : PEC-IT602B Data Warehousing and Data Mining

UPID : 006584

Time Allotted : 3 Hours *Full Marks :70*

The Figures in the margin indicate full marks.
Candidate are required to give their answers in their own words as far as practicable

Group-A (Very Short Answer Type Question)

1. Answer *any ten* of the following : [1 x 10 = 10]

(I) Explain the concept of prediction.

(II) How does autocorrelation impact time series analysis?

(III) What are some challenges in mining data streams?

(IV) Explain the difference between web content mining and web usage mining.

(V) What are some challenges in implementing distributed data mining?

(VI) Define data mining.

(VII) What is the significance of centroid-based clustering algorithms like k-means?

(VIII) What is the role of decomposition in time series analysis?

(IX) What is the importance of sampling in data stream mining?

(X) Discuss the ethical considerations in web mining.

(XI) What is the significance of modulation in communication systems?

(XII) Evaluate the challenges associated with data integration in data warehousing.

Group-B (Short Answer Type Question)

Answer *any three* of the following : [5 x 3 = 15]

2. Discuss the challenges associated with mining time series data and how they can be addressed. [5]
3. What are data streams, and how do they differ from static datasets in data mining? [5]
4. Explain the significance of mining the web page layout structure in web mining. [5]
5. How does graph mining contribute to extracting insights from interconnected data structures? Explain with a Neat Diagram. [5]
6. Discuss the significance of temporal-based frequent patterns in analyzing time-series data. [5]

Group-C (Long Answer Type Question)

Answer *any three* of the following : [15 x 3 = 45]

7. (a) What are some recent advancements in distributed warehousing technologies, and how do they impact data mining operations? [5]
   (b) Discuss the role of ensemble learning methods in addressing the class imbalance problem. [5]
   (c) How does graph mining contribute to anomaly detection in network data? [5]
8. (a) Illustrate how data mining techniques can be applied in retail to improve sales and customer satisfaction. [5]
   (b) Explain the significance of scalable methods in data mining and provide examples of scalable algorithms. [5]
   (c) Discuss the concept of correlation analysis in data mining and its applications. [5]
9. (a) Discuss the challenges associated with mining transactional patterns in large-scale datasets. [8]
   (b) Explain the concept of sequence mining and provide an example of its application. [7]
10. (a) Explain the difference between seasonal and non-seasonal patterns in time-related sequence data. [5]
    (b) Discuss the role of spectral analysis in detecting periodicity in time-related sequence data. [5]
    (c) How can mining time series data be used in predicting future trends or events? [5]
11. (a) Explain the data mining applications for retail industry. [6]
    (b) List the Issues to be considered during Data Integration. [5]

(c) Discuss about detecting data redundancy using correlation analysis. [4]

*** END OF PAPER ***