Code: AT19   **Subject: DATA WAREHOUSING AND DATA MINING**

AMIETE – IT (OLD SCHEME)

Time: 3 Hours | **OCTOBER 2012** | Max. Marks: 100

NOTE: There are 9 Questions in all.

- **Question 1 is compulsory and carries 20 marks. Answer to Q.1 must be written in the space provided for it in the answer book supplied and nowhere else.**
- **The answer sheet for the Q.1 will be collected by the invigilator after 45 Minutes of the commencement of the examination.**
- **Out of the remaining EIGHT Questions answer any FIVE Questions. Each question carries 16 marks.**
- **Any required data not explicitly given, may be suitably assumed and stated.**

Q.1   **Choose the correct or the best alternative in the following:**   **(2×10)**

a.   Retrieval of task relevant data in Data Mining is called

   (A) Data selection   **(B)** Pattern presentation
   (C) Data integration   **(D)** Task data

b.   Data mining is a synonym for

   (A) DBMS   **(B)** RDBMS
   (C) KDD   **(D)** Statistical Analysis

c.   Boosting applies to

   (A) Predictive data mining **(B)** Data warehousing
   (C) Statistical learning   **(D)** Learning

d.   Subset of data warehouse is

   (A) data mart   **(B)** database
   (C) data model   **(D)** meta data

e.   Which of the following is not an OLAP operation?

   (A) slice   **(B)** dice
   (C) roll-up   **(D)** union

f.   Tree pruning methods are used in

   (A) classification   **(B)** clustering
   (C) data cleaning.   **(D)** pattern recognition

g.   The normalized schema is called

(A) Star schema   **(B)** Snowflake schema
(C) Multidimensional schema   **(D)** Cube

h.   OLTP stands for

     (A) online transaction processing systems
     (B) offline transaction processing systems
     (C) online transaction systems
     (D) online table processing

i.   PCA is a technique used for

     (A) Mining patterns   **(B)** Reducing data.
     (C) Integrating data   **(D)** Cleaning data

j.   Metadata is

     (A) Data out of main data.
     (B) Data about data.
     (C) Separating primary & secondary data.
     (D) Partitioning of data.

Answer any FIVE Questions out of EIGHT Questions.
Each question carries 16 marks.

Q.2   a.   Define and differentiate between:
       (i) Executive information system & Decision support system.
       (ii) Primitive data and Derived data   **(4×2)**

     b.   List & explain different Data Warehouse tools that are used to populate
       and refresh data.   **(8)**

Q.3   a.   Describe the three tier architecture of data warehouse.   **(8)**

     b.   Distinguish between a snowflake schema and a star schema.   **(8)**

Q.4   a.   Explain the differences between operational database system and data
       warehouse.   **(8)**

     b.   Explain update-driven approach and query-driven approach for integrating
       multiple heterogeneous information sources.   **(8)**

Q.5   a.   Define data cleaning. How do you take care of missing values & noise in
       data cleaning?   **(8)**

     b.   Explain the different strategies for data reduction & data integration.   **(8)**

Q.6   a.   Explain the following OLAP operations with an example of each:
       (i)   Pivot
       (ii)   Slice and Dice.   **(4×2)**

      b.   Differentiate between Characterisation and Clustering.   **(8)**

Q.7   a.   What is external / unstructured data? Explain how such data is stored in a data warehouse.   **(8)**

      b.   What is the importance of metadata? Explain various components of Metadata.   **(8)**

Q.8   a.   Explain the significance of feedback loop for the success of data warehouse implementation.   **(8)**

      b.   What is EIS? List some of the typical uses of EIS.   **(8)**

Q.9   a.   Explain one data mining issue that, may have a strong impact on the market and on the society.   **(8)**

      b.   Explain data migration methodology with the help of a block diagram.   **(8)**