

Registration No:

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

Total Number of Pages: 02

B.Tech
PCS5H002

5th Semester Regular Examination 2017-18
Datamining & Data Warehousing

BRANCH: CSE

Time: 3 Hours

Max Marks: 100

Q.CODE: B307

Answer Question No.1 and 2 which are compulsory and any four from the rest.

The figures in the right hand margin indicate marks.

Q1 Answer the following questions: multiple type or dash fill up type (2 x 10)

- a) Which of the following is the most important when deciding on the data structure of a data mart? (210)
- (a) XML data exchange standards
 - (b) Data access tools to be used
 - (c) Metadata naming conventions
 - (d) Extract, Transform, and Load (ETL) tool to be used
- b) The process of removing the deficiencies and loopholes in the data is called as (210)
- (a) Aggregation of data
 - (b) Extracting of data
 - (c) Cleaning up of data.
 - (d) Loading of data
 - (e) Compression of data.
- c) Which one manages both current and historic transactions? (210)
- (a) OLTP
 - (b) OLAP
 - (c) Spread sheet
 - (d) XML
- d) Which of the following is the collection of data objects that are similar to one another within the same group? (210)
- (a) Partitioning
 - (b) Grid
 - (c) Cluster
 - (d) Table
 - (e) Data source.
- e) Which of the following process includes data cleaning, data integration, data selection, data transformation, data mining, pattern evolution and knowledge presentation? (210)
- (a) KDD process
 - (b) ETL process
 - (c) KTL process
 - (d) MDX process
 - (e) None of the above.
- f) Data mining application domains are (210)
- (a) Biomedical
 - (b) DNA data analysis
 - (c) Financial data analysis
 - (d) Retail industry and telecommunication industry
 - (e) All (a), (b), (c) and (d) above.
- g) Which of the following is not an ETL tool? (210)
- (a) Informatica
 - (b) Oracle warehouse builder
 - (c) Datastage
 - (d) Visual studio
 - (e) DT/studio.

- h) Which of the following is/are the Data mining tasks?
 (a) Regression
 (b) Classification
 (c) Clustering
 (d) inference of associative rules
 (e) All (a), (b), (c) and (d) above.
- i) Which of the following should not be considered for each dimension attribute?
 (a) Attribute name
 (b) Rapid changing dimension policy
 (c) Attribute definition
 (d) Sample data
 (e) Cardinality.
- j) Which of the following is the collection of data objects that are similar to one another within the same group?
 (a) Partitioning
 (b) Grid
 (c) Cluster
 (d) Table
 (e) Data source.

Q2 Answer the following questions: Short answer type (2 x 10)

- a) How is a data warehouse differ from a database ?
 b) Distinguish the feature between OLAP & OLTP .
 c) List data warehouse backend tools and its utilities and their functions.
 d) What is business intelligence ?
 e) What do you mean by neural clustering?
 f) Mention the utility of knowledge base.
 g) What is the drawback of using separate set of samples to evaluate pruning.
 h) List any two software tools associated with data mining and highlight their features.
 i) What are the steps involved in KDD process?
 j) Define meta data.

- Q3 a) Describe the architecture and implementation of data warehouse. (10)**
b) Briefly explain the basic dimensional modeling techniques. (5)

- Q4 a) Explain the algorithm for constructing a decisions tree from training samples. (10)**
b) Describe the K-Mean clustering algorithm. (5)

- Q5 a) What do you mean by data mining functionality ? Explain with suitable examples. (10)**
b) Explain OLAP operations in Multidimensional Data Model. (5)

- Q6 a) Explain the classification of major clustering methods. (10)**
b) Explain briefly about various steps of Data Mining process. (5)

- Q7 a) What is the role of data mining in spatial database ? (10)**
b) Detail on Data Warehouse meta data. (5)

- Q8 a) Explain in details about text mining applications. (10)**
b) How is web usage mining different from web structure mining and web content mining ? (5)

- Q9 a) Write short note on : (10)**
 i. Issues regarding classification and prediction.
 ii. Outlier Analysis.
b) Discuss about social impacts and various trends in Data Mining . (5)