

KADI SARVA VISHWAVIDYALAYA
B.E. SEMESTER 7TH EXAMINATION OCT - NOV 2016

SUBJECT CODE: IT – 701

SUBJECT NAME: BIG DATA ANALYTICS

DATE: 08/11/2016

TIME: 10:30 am to 01:30 pm

TOTAL MARKS: 70

- Instructions: 1) All questions are **compulsory**.
2) Figures to the **right** indicate full marks.
3) Indicate **clearly**, the options you attempt along with its respective question number.
4) Use the last page of main supplementary for **rough work**.

Section – 1

Marks

- Q.1 (A)** Consider the database below. What association rules can be found in this set using **Apriory Algorithm**, if the minimum support is 3 and the minimum confidence is 70%. **[5]**

TID List of item IDs

T1: 1,3,7

T2: 2,3,7

T3: 1,2,3

T4: 2,3

T5: 2,3,4,5

T6: 2,3

T7: 1,2,3,4,6

T8: 2,3,4,6

T9: 1

T10: 1,3

- Q.1 (B)** Consider the database given in Q-1(C) as above. What association rules can be found in this set using **FP – Growth Algorithm**, if the minimum support 3 and the minimum Confidence 70%. **[5]**

- Q.1 (C)** Explain Frequent Itemsets, Closed Itemsets, and Association Rules **[5]**

OR

- Q.1 (C)** Explain Enterprise Warehouse, Data Mart and Virtual Warehouse. **[5]**

- Q.2 (A)** Explain major steps for Hadoop Implementation. **[5]**

- Q.2 (B)** Differentiate the Big Data and traditional Enterprise Relational Data. **[5]**

OR

- Q.2 (A)** Explain Map-Reduce programming model with example of word count. **[5]**

- Q.2 (B)** Explain various components of Hadoop ecosystem. **[5]**

- Q.3 (A)** What is application of concept hierarchy? Draw concept hierarchy for Gujarat State – District and City wise and for time (year, quarter, month, week, day). **[5]**

- Q.3 (B)** What Is a Decision Tree? Explain with Examples, Advantages & Role in Enterprise. **[5]**

OR

- Q.3 (A)** Explain Supervise and Un Supervise learning with example. **[5]**

- Q.3 (B)** What is prediction? Compare Classification and Prediction according to different criteria. **[5]**

P.T.O.

Section – 2

Q.4 (A) Explain usage of data warehouse in the field of: Information processing, Analytical processing, and Data mining. [5]

Q.4 (B) Draw and explain Three – Tier Data Warehouse Architecture. [5]

Q.4 (C) A data warehouse can be modeled by either a star schema or a snowflake schema. Briefly describe the similarities and the differences of the two models, and then analyze their advantages and disadvantages with regard to one another. Give your opinion of which might be more empirically useful and state the reasons behind your answer. [5]

OR

Q.4 (C) Suppose that a data warehouse for KSV University consists of the following four dimensions: student, course, semester, and instructor, and two measures count and avg-grade. When at the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the avg grade measure stores the actual course grade of the student. At higher conceptual levels, avg grade stores the average grade for the given combination. Draw a snowflake schema diagram for the data warehouse. [5]

Q.5 (A) In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem. [5]

Q.5 (B) Explain typical OLAP operations on multidimensional data. [5]

OR

Q.5 (A) Explain Data Warehouse in terms of following: Subject – Oriented, Integrated, Time – Variant and Nonvolatile. [5]

Q.5 (B) What is Sampling? Explain different Sampling techniques. [5]

Q.6 (A) Use the two methods below to normalize the following group of data: [5]

2000, 3000, 4000, 6000, 10000

(a) Use min-max normalization to transform the value 5000 onto the range [0.0; 1.0].

(b) Use z-score normalization to transform the value 6000, where the standard deviation is 4500.

Q.6 (B) Explain Data Reduction techniques. [5]

OR

Q.6 (A) Explain major issues in Data Mining. [5]

Q.6 (B) What is noise? Explain different data smoothing techniques. [5]

KADI SARVA VISHWAVIDHYALAYA

B.E. SEMESTER – VII

EXAMINATION : (NOV – 2015)

SUBJECT CODE : IT – 701

SUBJECT NAME : BIG DATA ANALYTICS

DATE : 20.11.15

TIME : 10:30 to 1:30

TOTAL MARKS : 70

Instructions :

1. Answer each section in separate answer sheet.
2. Use of scientific calculator is permitted.
3. All questions are compulsory.
4. Indicate clearly, the options you attempted along with its respective question numbers.
5. Use the last-page of main supplementary for rough work.

SECTION - 1

Q-1: (All Compulsory)

- (A) Define Data Mining with the process of knowledge discovery. [5]
- (B) Define Data cleaning. Explain how to handle missing values during data cleaning? [5]
- (C) Explain Data Cube with all OLAP Operations in brief. [5]

OR

- (C) Compare Big Data and traditional enterprise relational data. [5]

Q-2: Answer the following questions.

- (A) Explain the 'Star' and 'Snowflake' schemas of data warehouse. [5]
- (B) What is Measures? List and explain types of measures. [5]

OR

- (A) Explain the information retrieval methods used in text mining. [5]
- (B) Explain Spatial, Legacy and Multimedia Database. [5]

Q-3: Answer the following questions.

- (A) Why naïve Bayesian classification is called "naïve"? Briefly outline the major ideas of naïve Bayesian classification. [5]
- (B) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order): [5]
13, 15, 16, 16, 19, 20, 23, 29, 35, 41, 44, 53, 62, 69, 72
(i) Use min-max normalization to transform the value 45 for age onto the range [0:0, 1:0]
(ii) Use z-score normalization to transform the value 45 for age, where the standard deviation of age is 20.64 years.

OR

- (A) State the Apriori Terminology. Generate large itemsets and association rules using Apriori Algorithm on the following data set with minimum support value and minimum Confidence value set as 50% and 75% respectively. [5]

TID	Items Purchased
T101	Cheese, Milk, Cookies
T102	Butter, Milk, Bread
T103	Cheese, Butter, Milk, Bread
T104	Butter, Bread

- (B) With the help of a neat diagram explain the 3-tier architecture of a data warehouse. [5]

SECTION - 2

Q-1: (All Compulsory)

- (A) Discuss the application of data warehousing and data mining in government sector. [5]
(B) Define cluster and explain any one (K-means or k-medoids) clustering algorithm in detail. [5]
(C) Explain Linear Regression and Non-linear Regression techniques of prediction. [5]

OR

- (C) What is Noisy data? List and describe data smoothing techniques. [5]

Q-2: Answer the following questions.

- (A) Compare OLTP & OLAP systems. [5]
(B) Explain Rule-based Classification in brief. [5]

OR

- (A) Explain Market Basket Analysis with its use and Association Rules in brief. [5]
(B) Explain the Classification by Decision Tree Induction Algorithm. [5]

Q-3: Answer the following questions.

- (A) Define Big data and discuss Challenges of Big Data. [5]
(B) Discuss importance of Map-reduce in data analytics. [5]

OR

- (A) Explain Hadoop architecture in detail. [5]
(B) Discuss map reduce working with example. [5]