E/Sem-VIII-(CBSGS)/CMPN/Data Warehouse & Mining/M-18

Q. P. Code: 25961

Time: 03 Hours **Marks: 80**

Note: 1. Question 1 is compulsory

2. Answer any three out of remaining questions.

Q1 A) i. Design star & snowflake schema for "Hotel Occupancy" considering dimensions like Time, Hotel, Room, etc. [10]

ii. Calculate the maximum number of base fact table records for the values given below:

Time period: 5 years

Hotels: 150

Rooms: 750 rooms in each Hotel (about 400 occupied in each hotel daily).

B) Explain Data mining as a step in KDD. Give the architecture of typical data mining System. [10]

Q2 A) The college wants to record the marks for the courses completed by students using the dimensions: a) Course, b) Student, c) Time & a measure d) Aggregate marks. Create a Cube and describe following OLAP operations: [10]

i) Rollup ii) Drill down iii) Slice iv) Dice v) Pivot.

B) A simple example from the stock market involving only discrete ranges has profit as categorical attribute, with values {up, down} and the training data is: [10]

Age	Competition	Type	Profit
Old	Yes	Software	Down
Old	No	Software	Down
Old	No	Hardware	Down
Mid	Yes	Software	Down
Mid	Yes	Hardware	Down
Mid	No	Hardware	Up
Mid	No	Software	Up
New	Yes	Software	Up
New	No	Hardware	Up
New	No	Software	Up

Apply decision tree algorithm and show the generated rules.

Q3 A) Why naive Bayesian classification is called "naive"? Briefly outline the major ideas of Naive Bayesian classification. [10]

B) Discuss different steps involved in Data Pre-processing [10]

Q4 A) Explain ETL of data warehousing in detail. [10]

B) Find clusters using k-means clustering algorithm if we have several objects (4 types of medicines) and each object have two attributes or features as shown in the table below. The goal is to group these objects into k=2 group of medicine [10]

based on the two features (pH and weight index).

Object	Attribute 1(X) Weight Index	Attribute 2 (Y) pH
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

Q5 A) Explain Data Warehouse Architecture in detail. [10]

B) A database has five transactions. Let minimum support = 30% and minimum Confidence = 70% [10]

i. Find all frequent patterns using Apriori Algorithm.

ii. List strong association rules.

Transaction _Id	Items
A	1,3,4,6
B	2,3,5,7
C	1,2,3,5,8
D	2,5,9,10
E	1,4

Q6 Write short note on the following (Answer any **FOUR**) [20]

a) Data warehouse design strategies

b) Applications of Data Mining

c) Role of metadata

d) Multidimensional and multilevel association mining

e) Hierarchical clustering

****

F765B4AF9629D2B23406B13A0E2E0BA5

BE/CMPN/CBSGS/SEM VIII/HMI/17-18

961

Q.P. Code: 27190

Time: 3 Hours **(Total Marks 80)**

Question no 1 is compulsory
solve any 3 from Q2 to Q6
Indicate your answers with neat sketch wherever necessary

Q1 . ATTEMPT ANY FOUR **20**

a) What do you mean by response time? Explain.
b) Explain three levels of processing in detail.
c) State and explain different interview techniques.
d) Explain steps in constructing persona.
e) How one can create a dialogue with interface user? Explain with appropriate example.

Q2 a) What are various type of windows? Explain. **10**
b) What are three levels of users? Explain how to accommodate them in user interface. **10**

Q3 a) What are various factors to be considered for User Interface Design? Also give suitable example for the same. **10**
b) Differentiate between Qualitative and Quantitative Research **10**

Q4 a) What are various methodologies adopted for Feedback and Guidance? Explain in detail. **10**
b) Explain Various Menus in Human Machine Interface? **10**

Q5 a) What do you mean by Keyboard Accelerators? Explain **10**
b) Explain Goal Directed Design Process in Detail. **10**

Q6 Write Short Note on following **20**

a) Device Based Control
b) Screen Based Control
c) Statistical Graphics
d) Graphics Icons and Images

Sem VIII (CBSGS) / EXTC / Parallel & Distributed System / M-18.

Q. P. Code: 35833

3Hrs **80 Marks**

Note: 1) Question no 1 is compulsory

2) Solve any three from remaining

Q1 a) Explain Flynn's classification scheme (05)

b) Explain Data mapping and memory in array processor (05)

c) Explain desirable features of global scheduling algorithm (05)

d) Explain Berkeley physical clock algorithm (05)

Q2 a) Explain different types of Hazards in Parallel System (10)

b) Explain Ricart-Agrawala algorithm for Mutual Exclusion (10)

Q3 a) Give an example that can be solved effectively with SIMD architecture (10)

b) What are the different Architectural Model of Distributed System?

Explain with suitable diagram (10)

Q4 a) Explain Hadoop Distributed File System (HDFS). (10)

b) Explain Software models supported by the distributed system (10)

Q5 a) What is Remote Procedure Call? Explain the working of RPC in detail. (10)

b) What are different data centric consistency model? (10)

Q6 a) Explain different load estimation policies and process transfer policies used by Load balancing algorithm (10)

b) Explain Bully Election Algorithm (10)

----------x----------

Sem-VIII (CBSGS) / EMPN / Elective - Machine Learning /
May -2018.

Q.P. Code :16172

[Time: 3 Hours] **[Marks:80]**

Please check whether you have got the right question paper.

N.B:
1. Question no 1 is compulsory
2. Attempt any three questions out of remaining five questions
3. Assume any suitable data wherever required but justify the same.

Q1 a. Define Machine Learning? Briefly explain the types of learning. **05**

b. What is independent component analysis? **05**

c. What are the issues in decision tree induction? **05**

d. What are the requirements of clustering algorithms? **05**

Q2 a. The values of independent variable x and dependent value y are given below: **10**

X	Y
0	2
1	3
2	5
3	4
4	6

Find the least square regression line y=ax+b. Estimate the value of y when x is 10.

b. What are the steps in designing a machine learning problem? Explain with the checkers problem. **10**

Q3 a. For a SunBurn dataset given below, construct a decision tree **10**

Name	Hair	Height	Weight	Location	Class
Sunita	Blonde	Average	Light	No	Yes
Anita	Blonde	Tall	Average	Yes	No
Kavita	Brown	Short	Average	Yes	No
Sushma	Blonde	Short	Average	No	Yes
Xavier	Red	Average	Heavy	No	Yes
Balaji	Brown	Tall	Heavy	No	No
Ramesh	Brown	Average	Heavy	No	No
Swetha	Blonde	Short	Light	Yes	No

b. What is the goal of the Support Vector Machine (SVM)? How to compute the margin? **10**

Q4 a. For the given set of points identify clusters using complete link and average link using agglomerative clustering. **10**

	A	B
P1	1	1
P2	1.5	1.5
P3	5	5
P4	3	4
P5	4	4
P6	3	3.5

b. What is the role of radial basis function in separating nonlinear patterns. **10**

Q.5 a. Use Principal Component analysis (PCA) to arrive at the transformed matrix for the given matrix A. **10**

$A^T = \begin{matrix} 2 & 1 & 0 & -1 \\ 4 & 3 & 1 & 0.5 \end{matrix}$

b. What are the elements of reinforcement learning? **10**

Q.6 Write short notes on any two **20**

a. Logistic regression
b. Back propagation algorithm
c. Issues in Machine Learning

Sem- XIII -(CBSGS) / EMPN / Elective - Big Data Analytics / May-2018.

Q.P.Code: 40458

(3 Hours) [Total Marks 80]

i. **Q. 1. is Compulsory.**
ii. **Attempt any three from the remaining.**
iii. **Assume suitable data.**

Q. 1 **(a)** Explain what characteristics of Social Networks make it Big Data. **(5)**

(b) What do you mean by Jaccard Similarity? Illustrate with an example. Describe any two applications that can use Jaccard Similarity. **(5)**

(c) Define concept of a **Link Farm** using a diagram. How does it lead to Link Spam? **(5)**

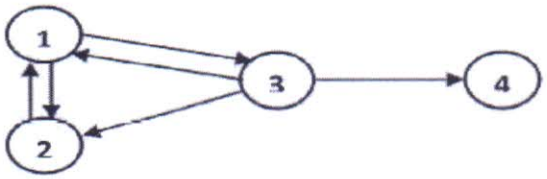
(d) What are the challenges of querying on large Data Streams? **(5)**

Q. 2 **(a)** What do you understand by BASE properties in NOSQL Database? Explain in detail any one NOSQL architecture pattern. Identify two applications that can use this pattern. **(10)**

(b) Write Map Reduce Pseudocode to multiply two matrices. Illustrate the procedure on the following matrices. Clearly show all the steps. **(10)**

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$$

Q. 3 **(a)** For the graph given below show the page ranks of all the nodes after running the PageRank algorithm for two iterations with teleportation factor with Beta (β) value = 0.8 **(10)**



(b) Give two applications for counting the number of 1's in a long stream of binary values. Using a stream of binary digits, illustrate how the DGIM algorithm will find the number of 1's. **(10)**

Q. 4 (a) What do you mean by the Hadoop Ecosystem? Describe any three components of a typical Hadoop Ecosystem. **(10)**

(b) Explain the following concepts with respect to world wide web **(10)**

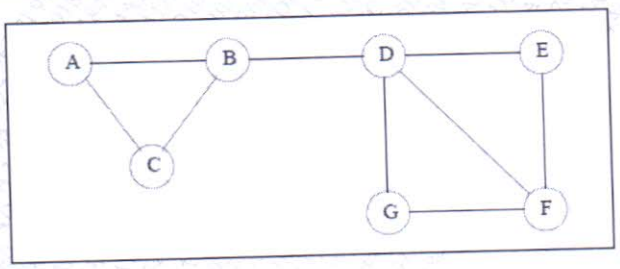
A. Topic Specific Page Rank
B. Bowtie structure of the Web

Q. 5 (a) Explain the design of a recommender system used to recommend movies to users. The recommender system should use Collaborative filtering. **(10)**

(b) Explain the CURE algorithm for clustering large datasets. Please illustrate the algorithm using appropriate figures. **(10)**

Q. 6 (a) Explain the SON algorithm for Frequent Pattern mining. Illustrate how Map Reduce can be used for implementing this algorithm **(10)**

(b) What is a "Community" in a Social Network Graph? For the following graph show how the Girvan Newman algorithm finds the different communities. **(10)**



****