Q.P. Code: 25960

Time: 03 Hours **Marks: 80**

Note: 1. Question 1 is compulsory

2. Answer any three out of remaining questions.

Q1 A) Consider following dimensions for a Supermarket chain: Product, Store, Time and Promotion. With respect to this business scenario, answer the following questions. Clearly state any reasonable assumptions you make. [10]

(a) Design an information package diagram for this business scenario.

(b) Design a snowflake schema for the data warehouse, clearly depicting the fact table(s), Dimension table(s), their attributes and measures.

B) Consider the 5 transactions given below. If minimum support is 30% and minimum confidence is 80%, determine the frequent itemsets and association rules using Apriori algorithm. [10]

Transaction	Items
T1	Milk, Jam, Butter
T2	Milk, Butter
T3	Milk, Cheese, Butter
T4	Biscuit, Milk,
T5	Biscuit, Cheese

Q2 A) Consider a Data Warehouse for a sport manufacturing company storing sales details of various sports equipments sold, and the time of the sale. Using this example describe the following OLAP operations: [10]
(i) Slice (ii) Dice (iii) Rollup (iv) Drill Down (v) Pivot

B) What is data mining? Describe the steps involved in the data mining when viewed as a process of knowledge discovery. Present an example where data mining is crucial to success of business. [10]

Q3 A) What is Dimension Modeling? What is slowly changing dimensions? How this problem is solved? Give example. [10]

B) Given is the training data for height classification, classify the **tuple t= <Arvish, M, 1.97 >** using Bayesian classification. [10]

Name	Gender	Height	Output
Reena	F	1.6 m	Short
Mahesh	M	2 m	Tall
Tina	F	1.9 m	Medium
Meeta	F	1.88 m	Medium

Siya	F	1.7 m	Short
Vikram	M	1.85 m	Medium
Lakshmi	F	1.6 m	Short
Andrew	M	1.7 m	Short
Henry	M	2.2 m	Tall
Akhil	M	2.1 m	Tall
Lata	F	1.8 m	Medium
Siraj	M	1.95 m	Medium
Rita	F	1.9 m	Medium
Kriti	F	1.8 m	Medium
Srishti	F	1.75 m	Medium

Q4 A) Differentiate between top-down and bottom-up approaches for building data warehouse. Discuss the merits and limitations of each approach. Also explain the practical approach for designing a data warehouse. [10]

B) What is clustering? Explain K means clustering algorithm. [10]
Suppose the data for clustering is {2, 4, 10, 12, 3, 20, 30, 11, 25, 5, 36, 41, 14}. Assuming number of clusters to be 2 i.e. K = 2, cluster the given data using above algorithm.

Q5 A) Describe different steps of ETL (Extraction, Transformation and Loading) cycle in Data Warehousing for a pharmaceutical company. [10]

B) What is Web Mining? Explain Web Usage Mining. [10]

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

a) Hierarchical Clustering Algorithms

b) Metadata in Data Warehouse

c) Decision tree Classification Model

d) Snapshot and Transaction tables

e) Data Exploration

****

CF2EE518135A3F04241DECDA8318E12D

CMPN/CBSGS/SEM VIII/Paper / Subject Code: 52707 / Human Machine Interaction / MAY 2019

Max.Marks:80

Time: 3 hours

Note: 1. Question 1 is compulsory.
2. Attempt any 3 from Q2 to Q6.
3. Indicate your answer with various sketches whenever necessary.

Q1. Attempt any **four**. [20]
(a) List pros and cons of any one modern device in design of a tutor for kids.
(b) List techniques in qualitative research.
(c) Differentiate between direct and indirect manipulation.
(d) Explain goal directed design in brief.
(e) What are keyboard accelerators? Explain.

Q2. (a) Provide all factors of UI design. Give an example for incorporating innovative technologies. [10]
(b) Explain in details Gestalts principal. [10]

Q3. (a) Give brief description of GUI and web papges. [10]
(b) Explain seven stages of action and three levels of processing. [10]

Q4 (a) Explain six behavioral patterns in details. [10]
(b) Differentiate between quantitative and qualitative research in knowing the users. [10]

Q5 (a) State and explain principles of Gestalts theory. Give example. [10]
(b) Provide suitable analysis and Interface design for state road transportation system. [10]

Q6 Write Short notes on following **(Any Four)**. [20]
(a) Statistical Graphics
(b) Guidance and Feedback
(c) Interview Questions
(d) Goal directed Design
(e) Device based control
(f) Usability Design Principles

****

Sem-VIII | Paper / Subject Code: 52708 / Parallel and Distributed Systems | May-2019

PN | CBSGS.

(3 hrs) Marks: 80

N.B. (1) Question one is Compulsory.
(2) Attempt any 3 questions out of the remaining.
(3) Assume suitable data if required.

Q1. a) What are various issues of distributed system? 05

b) Suppose through experimentation it was verified that 70% of execution was spent on parallelizable execution. What are the maximum speedup and efficiency those can be achieved with 8 processors? 05

c) Justify how Ricart-Agrawala's algorithm optimized the Message overhead in achieving mutual exclusion. 05

d) Give examples for the following message communication models 05

- Transient Synchronous
- Response based synchronous communication
- Transient asynchronous
- Persistent Asynchronous
- Receipt based communications

Q2. a) Brief the different load estimation policies and process transfer policies used by Load balancing algorithm. 10

b) Discuss the Structural and Data hazards in Pipeline architecture. Discuss any one technique to control / mitigate them in detail. 10

Q3. a) Design and analyze 3-stage pipeline operations executing the following task: $Xn + Yn * Zn$, for $n = 1, 2, 3, \ldots, 7$. 10

b) Describe any one method of Logical Clock synchronization. 10

Q4 a) Clearly explain how Monotonic Read consistency model is different from Read your Write Consistency model. Support your answer with suitable example application scenarios where each of them can be distinctly used. 10

b) Discuss the need for process migration and the role of resource to process and process to resource binding in process migration 10

Q5 a) Apply quicksort parallel algorithm for the following example: 16, 08, 33, 45, 25, 19, 53, 06 10

b) Differentiate between Distributed OS, Network OS and Middleware based OS 10

Q6 Write a note on **any two** of the following 20

a) Hadoop Distributed File System
b) Systolic Architecture
c) RPC and RMI

****

BE|SEM VIII|CMPN| **Paper / Subject Code: 52701 / Elective- III 1) Machine Learning**|MAY 2019

Time: 03 Hours **Marks: 80**

Note: 1. Question 1 is compulsory
2. Answer any three out of remaining five questions.
3. Assume any suitable data wherever required and justify the same.

Q1 a) Define Machine Learning (ML) Briefly explain the types of learning. [5]

b) "Entropy is a thermodynamic function used to measure the disorder of a system in Chemistry." How do you suitably clarify the concept of entropy in ML? [5]

c) State the principle of Occam's Razar. Which ML algorithm uses this principle? [5]

d) Explain Bayesian Belief Network with an example. [5]

Q2 a) Use the k-means clustering algorithm and Euclidean distance to cluster the following eight 8 examples into three clusters: [10]
A1= (2, 10), A2= (2, 5), A3= (8, 4), A4= (5, 8), A5= (7, 5), A6= (6, 4), A7= 1(1, 2), A8= (4, 9). Find the new centroid at every new point entry into the cluster group. Assume initial cluster centers as A1, A4 and A7.

b) Compare and contrast Linear and Logistic regressions with respect to their mechanisms of prediction. [10]

Q3 a) Find predicted value of Y for one epoch and RMSE using Linear regression. [10]

X	Y-Actual
2	1
3	3
4	6
5	9
6	11
7	13
8	15
9	17
10	20

b) Find the new revised *theta* for the given problem using Expectation -Maximization Algorithm for one epoch. [10]

1	H	T	T	T	H	H	T	H	T	H
2	H	H	H	H	H	T	H	H	H	H
3	H	T	H	H	H	H	H	T	H	H
4	H	T	H	T	T	T	H	H	T	T
5	T	H	H	H	T	H	H	H	T	H

$\Theta_A = 0.6$ and $\Theta_B = 0.5$

BE/SEM VIII/CMPN/ **Paper / Subject Code: 52701 / Elective- III 1) Machine Learning**/ MAY 2019

Time: 03 Hours **Marks: 80**

Note: 1. Question 1 is compulsory
2. Answer any three out of remaining five questions.
3. Assume any suitable data wherever required and justify the same.

Q1 a) Define Machine Learning (ML) Briefly explain the types of learning. [5]

b) "Entropy is a thermodynamic function used to measure the disorder of a system in Chemistry." How do you suitably clarify the concept of entropy in ML? [5]

c) State the principle of Occam's Razar. Which ML algorithm uses this principle? [5]

d) Explain Bayesian Belief Network with an example. [5]

Q2 a) Use the k-means clustering algorithm and Euclidean distance to cluster the following eight 8 examples into three clusters: [10]
A1= (2, 10), A2= (2, 5), A3= (8, 4), A4= (5, 8), A5= (7, 5), A6= (6, 4), A7= 1(1, 2), A8= (4, 9). Find the new centroid at every new point entry into the cluster group. Assume initial cluster centers as A1, A4 and A7.

b) Compare and contrast Linear and Logistic regressions with respect to their mechanisms of prediction. [10]

Q3 a) Find predicted value of Y for one epoch and RMSE using Linear regression. [10]

X	Y-Actual
2	1
3	3
4	6
5	9
6	11
7	13
8	15
9	17
10	20

b) Find the new revised *theta* for the given problem using Expectation -Maximization Algorithm for one epoch. [10]

1	H	T	T	T	H	H	T	H	T	H
2	H	H	H	H	H	T	H	H	H	H
3	H	T	H	H	H	H	H	T	H	H
4	H	T	H	T	T	T	H	H	T	T
5	T	H	H	H	T	H	H	H	T	H

$\Theta_A = 0.6$ and $\Theta_B = 0.5$

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

N.B. Question No: 1 is Compulsory
Attempt any three from the remaining
Assume suitable data wherever necessary

1 a Find Manhattan distance for the points X1= (1,2,2) , X2 = (2,5,3) **5**
b How finding plagiarism in documents is a nearest neighbor problem. **5**
c Draw and Explain Bow-tie structure of web. **5**
d How big data problems are handled by Hadoop system. **5**

2 a Explain how Hadoop goals are covered in hadoop distributed file system. **10**
b Write pseudo code for Matrix vector Multiplication by MapReduce. Illustrate with an example showing all the steps. **10**

3 a The snapshot of 10 transactions is given below for online shopping that generates big data. Threshold value = 4 and Hash function= (i*j) mod 10 **10**

T1 = {1, 2, 3} T2 = {2, 3, 4} T3 = {3, 4, 5}
T4 = {4, 5, 6} T5 = {1, 3, 5} T6 = {2, 4, 6}
T7 = {1, 3, 4} T8 = {2, 4, 5} T9 = {3, 4, 6} T10 = {1, 2, 4}

Find the frequent item sets purchased for such big data by using suitable algorithm. Analyse the memory requirements for it.
b Explain DGIM algorithm for counting ones in stream with example. **10**

4 a How recommendation is done based on properties of product? Explain with suitable example. **10**
b Explain how the CURE algorithm can be used to cluster big data sets. **10**

5 a What are the different architectural patterns in NoSQL? Explain Graph data store and Column Family Store patterns with relevant examples. **10**
b Explain Girvan-Newman algorithm to mine Social Graphs. **10**

6 a List down the steps in modified Page Rank Algorithm to avoid spider trap with one example. **10**
b Explain Park-Chen-Yu algorithm. How memory mapping is done in PCY. **10**

****