

(3 Hours)

[Total Marks : 80

Note: 1. Question No.1 is compulsory

2. Attempt any **Three** questions out of remaining questions

3. Assume suitable data wherever necessary and state them clearly

Q1 a) For a Super market chain, consider the following dimensions namely product, store, [10]
time and promotion. The schema contains a central fact table for sales.

- Design star schema for the above application.
- Calculate the maximum number of base fact table records for warehouse with the following values given below:
 - Time period – 5 years
 - Store – 300 stores reporting daily sales
 - Product – 40,000 products in each store (about 4000 sell in each store daily)

b) Discuss: [10]

- The steps in KDD process
- The architecture of a typical DM system

Q2 a) We would like to view sales data of a company with respect to three dimensions namely [10]
Location, Item and Time. Represent the sales data in the form of a 3-D data cube for the above and Perform Roll up, Drill down, Slice and Dice OLAP operations on the above data cube and Illustrate.

b) A simple example from the stock market involving only discrete ranges has profit as [10]
categorical attribute, with values {Up, Down} and the training data set is given below.

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) Illustrate the architecture of a typical DW system. Differentiate DW and Data Mart. [10]

b) Discuss different steps involved in Data Preprocessing. [10]

Q4 a) Discuss various OLAP Models. [10]

b) Explain K-Means clustering algorithm? Apply K-Means algorithms for the following [10]
data set with two clusters. Data Set = {1, 2, 6, 7, 8, 10, 15, 17, 20}

TURN OVER

Q5 a) Describe the steps of ETL process. [10]

b) Discuss Association Rule Mining and Apriori Algorithm. Apply AR Mining to find all frequent item sets and association rules for the following dataset: [10]

Minimum Support Count = 2

Minimum Confidence = 70%

Transaction ID	Items
100	1, 2, 5
200	2, 4
300	2, 3
400	1, 2, 4
500	1, 3
600	1, 3
700	1, 3, 2, 5
800	1, 3
900	1, 2, 3

Q6 Write Short notes on any four of the following: [20]

- Updates to Dimension tables
- Metrics for Evaluating Classifier Performance
- FP tree
- Multilevel & Multidimensional Association Rule
- Operational Vs. Decision Support System

N.B. : (1) Q 1 is **Compulsory**

(2) Solve any 3 from remaining Questions

(3) Please specify your answers with neat sketches and Example wherever necessary

(4) Assume any data if not specified

1. Solve any **five** of the following.

20

(a) What do you mean by response time?

(b) What do you mean by direct manipulation and indirect manipulation

(c) Discuss issues related to Long Term Memory and short Term Memory.

(d) What do you mean by keyboard accelerators?

(e) Explain importance of Text Messages with respect to communication with user.

(f) What are the three levels of processing? Explain

2 (A) Give Brief description on GUI versus Web Page.

10

(B) In the state of Maharashtra, Rationing Department want to provide self-help portal for its customers. The portal consists of Information about basic need supplies, Online application for ration cards, Schemes for Low income groups, Adding a name of family members, deletion of name, Complaints and other facilities. Being a Subject Matter Expert (SME) provide the detailed analysis and for the same provide the Interface that will be used by people in all Districts of Maharashtra

10

3. (A) What do you mean by Qualitative and Quantitative Research? State various interview technics?

10

(B) Provide suitable Analysis and Interface Design for State Road Transport corporations Information KIOSK that will be installed on major Bus stands of Indian state and it will have Local National and International Language (English) consider Hindi or Marathi as a local Language and provide detailed analysis.

10

4. (A) Provide all factors of Interface design? Provide innovative web application by integrating the technologies that are used in Interface design. 10
- (B) Provide a systematic design analysis for Municipal Corporation's Mobile App; that provides information about the wards, their ward office, corporators in the ward, Schools Hospitals in the ward and other information of the Municipal Office, your analysis should consist of all necessary interface guidelines. 10
5. (A) What do you mean by persona? Mention steps in constructing persona 10
- (B) What are various methodologies adopted for Feedback and guidance? Consider multimedia, File Download and Software Installation and state how we can create a dialog with user to communicate the Time factor for each one. 10
6. Write Short Notes on ANY FOUR 20
- (a) Human Centric Design
 - (b) Designs for accommodating Users
 - (c) Logo Design
 - (d) Menus in HMI
 - (e) Graphics Icons and Images
 - (f) Windows
-

N.B. : (1) Question No. 1 is **Compulsory**

(2) Attempt any **three** questions out of remaining **five** questions.

(3) Assume suitable data wherever required but justify the same.

(4) Assumption made should be clearly stated.

1. (a) Explain Data Flow computers with example. 5
(b) What is the basic task of scheduler? Define i) Latency, ii) Initiation Rate, iii) Stage Utilization and iv) Forbidden Latency. 5
(c) What are the different models of middleware? 5
(d) What are the issues in designing a distributed system? 5
2. (a) A machine is run on many applications and the instruction mix is collected. Loads/Store are 10%, Integer add/sub 15%, FP add/sub 50%, FP multiply/divide 5% and others 5% and branches 15%. The clock cycles consumed by these instructions are: Loads 2, Integer add/sub 1, FP add/sub 5, FP multiply/divide 20, others 1. Find which component of the architecture requires enhancement first. After incorporating the enhancement which makes clock cycles requirements as 2. Find the overall Speedup? 10
(b) What is SIMD Architecture? Explain with example SIMD Mesh Connected Architecture. 10
3. (a) What is an interlock? Explain the following three different classes of hazards : 10
(i) Control Hazards
(ii) Resource hazards
(iii) Operand hazards
(b) Explain a pipelined multiplication using Digit Products of Fixed Point Multiplication Pipeline. 10
4. (a) Explain the difference between Data Centric and Client Centric Consistency Models. Explain one model of each. 10
(b) Explain stream oriented communication with suitable example 10

Q.P. Code : 735400

2

5. (a) Explain the distributed algorithms for Mutual Exclusion? What are the advantages and disadvantages of it over centralized algorithms? **10**
(b) Write a Suzuki-Kasami's Broadcast Algorithm. Explain with example. **10**
6. (a) Compare Load sharing to task assignment and Load balancing strategies for scheduling processes in a distributed system. **10**
(b) What are the desirable features of good distributed file systems? **10**
Explain file sharing semantic of it.
-

MUPD16615 VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY 5/30/2016 9:45:39 AM MUPD16615 VIVEKANAND

N.B.:

1. Question No.1 is **compulsory**.
2. Attempt any **Three** questions out of remaining **Five** questions.
3. Figures to the right indicate full marks.
4. Assume any suitable data wherever required but justify the same.

- Q.1 a) What are the key tasks of Machine Learning? 5
- b) What are the key terminologies of Support Vector Machine? 5
- c) Explain in brief Linear Regression Technique. 5
- d) Explain in brief elements of Reinforcement Learning. 5
- Q.2 a) Explain the steps required for selecting the right machine learning algorithm. 8
- b) For the given data determine the entropy after classification using each attribute for classification separately and find which attribute is best as decision attribute for the root by finding information gain with respect to entropy of Temperature as reference attribute. 12

Sr. No.	Temperature	Wind	Humidity
1	Hot	Weak	High
2	Hot	strong	High
3	Mild	Weak	Normal
4	Cool	Strong	High
5	Cool	Weak	Normal
6	Mild	Strong	Normal
7	Mild	Weak	High
8	Hot	strong	High
9	Mild	Weak	Normal
10	Hot	Strong	Normal

- Q.3 a) Explain in detail Principal Component Analysis for Dimension Reduction 10
- b) Apply K-means algorithm on given data for k=3. Use $C_1(2)$, $C_2(16)$ and $C_3(38)$ as initial cluster centres. 10
- Data: 2, 4, 6, 3, 31, 12, 15, 16, 38, 35, 14, 21, 23, 25, 30
- Q.4 a) Explain in detail reinforcement technique Temporal Difference Learning. 10
- b) Using Bayesian classification and the given data classify the tuple(Rupesh, M, 1.73 m) 10

Attribute	Value	Count			Probability		
		Short	Medium	Tall	Short	Medium	Tall
Gender	M	1	2	3	1/4	2/7	3/4
	F	3	5	1	3/4	5/7	1/4
Height	(0, 1.6)	2	0	0	2/4	0	0
	(1.6, 1.7)	2	0	0	2/4	0	0
	(1.7, 1.8)	0	3	0	0	3/7	0
	(1.8, 1.9)	0	3	0	0	3/7	0
	(1.9, 2)	0	1	2	0	1/7	2/4
	(2, ∞)	0	0	2	0	0	2/4

TURN OVER

- Q.5 a) Apply Agglomerative clustering algorithm on given data and draw dendrogram. Show three clusters with its allocated points. Use single link method. 8

Adjacency matrix

	a	b	c	d	e	f
a	0	$\sqrt{2}$	$\sqrt{10}$	$\sqrt{17}$	$\sqrt{5}$	$\sqrt{20}$
b	$\sqrt{2}$	0	$\sqrt{8}$	3	1	$\sqrt{18}$
c	$\sqrt{10}$	$\sqrt{8}$	0	$\sqrt{5}$	$\sqrt{5}$	2
d	$\sqrt{17}$	3	$\sqrt{5}$	0	2	3
e	$\sqrt{5}$	1	$\sqrt{5}$	2	0	$\sqrt{13}$
f	$\sqrt{20}$	$\sqrt{18}$	2	3	$\sqrt{13}$	0

- b) Explain classification using Back Propagation algorithm with a suitable example. 12

- Q.6 Write detail notes on (any two) 20

- Quadratic Programming solution for finding maximum margin separation in Support Vector Machine.
 - Applications of Machine Learning algorithms.
 - Hidden Markov Model.
-

QP Code : 725400

(3 Hours)

[Total Marks :80

- N.B. : (1) Question no. 1 is **compulsory**
(2) Attempt any three from the remaining.
(3) Assume suitable data.

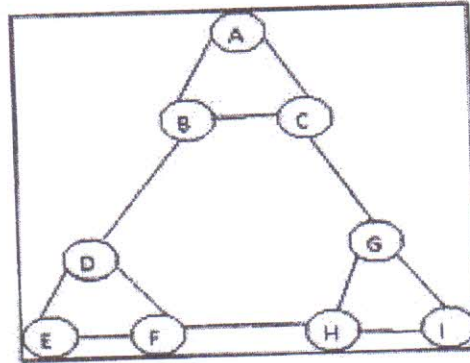
1. (a) What are the three Vs of Big Data? Give two examples of big data case studies. 5
Indicate which Vs are satisfied by these case studies.
- (b) Describe the operations of "shuffle" and "sort" in the Map reduce framework? Explain with the help of one example. 5
- (c) Through an example illustrate how triples can be used to optimally store and count pairs in a frequent itemset mining algorithm. 5
- (d) What is the motivation to count triangles in a social network graph? 5
Outline one algorithm for counting triangles briefly.
2. (a) What are the different data architecture patterns in NOSQL? Explain "Graph Store" and "Column Family Store" patterns with relevant examples. 10
- (b) Show Map Reduce implementation for the following two tasks using pseudocode. 10
 - (i) Join of two relations with an example
 - (ii) Multiplication of two matrices with one Map reduce step.
3. (a) Write a note on different distance measures that can be used to find similarity/dissimilarity between data points in a big data set. 10
- (b) Describe any two sampling techniques for big data with the help of examples. 10
4. (a) Using an example bit stream explain the working of the DGIM algorithm to count number of 1's (Ones) in a data stream. 10
- (b) Clearly explain how the CURE algorithm can be used to cluster big data sets. 10
5. (a) Define Content based recommendation systems. Using an example case study describe how it can be used to provide recommendations to users. 10
- (b) Let the adjacency matrix for a graph of four vertices $\{n_1 \text{ to } n_4\}$ be as follows: 10

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Calculate the authority and hub scores for this graph using the HITS algorithm with $k = 6$, and identify the best authority and hub nodes.

[TURN OVER]

6. (a) Explain clearly how the SON partition based algorithm helps to perform frequent itemset mining for large datasets. How does this algorithm avoid False Negatives? 10
 (b) For the graph given below use Clique percolation and find all communities 10



MUPD16615 VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY 5/12/2016 9:45:59 AM MUPD16615 VIVEKANAND