Time: 3hrs **[Total Marks:80]**

N.B. : (1) Question No 1 is Compulsory.
(2) Attempt any three questions out of the remaining five.
(3) Assume suitable data, if required and state it clearly.

Q1 Attempt any **FOUR** from the following [20]

A Explain how to choose the right algorithm for machine learning application.

B Explain Linear Discriminant Analysis.

C Explain any five performance measures along with example.

D Differentiate between Logistic regression and Support vector machine.

E Explain the following Receiver operating characteristics curve and Area under curve.

Q2 A Explain clustering with minimal spanning tree with reference to Graph based clustering. [10]

B Explain the terms overfitting, underfitting, bias & variance tradeoff w.r.t. Machine Learning. [10]

Q3 A Explain the concept of regression and enlist its types. A clinical trial gave the data for BMI and Cholesterol level for 10 Patients as shown in table below. Identify the machine learning method used to solve the above problem and predict the likely value of Cholesterol level for someone who has BMI of 27. [10]

BMI	17	21	24	28	14	16	19	22	15	18
Cholesterol	140	189	210	240	130	100	135	166	130	170

B Explain the necessity of cross validation in Machine learning applications and K-fold cross validation in detail. [10]

Q4 A Explain support vector machine as a constrained optimization problem. [10]

B Explain the concept of decision tree. Consider the dataset given in a table below. The dataset has 3 features as Past Trend, Open interest, Trading volume and one class label as Return. Compute the Gini Index for all features and specify which node will be chosen as a root node in decision tree. [10]

Past Trend	Open Interest	Trading Volume	Return
Positive	Low	High	Up
Negative	High	Low	Down
Positive	Low	High	Up
Positive	High	High	Up
Negative	Low	High	Down
Positive	Low	Low	Down
Negative	High	High	Down
Negative	Low	High	Down
Positive	Low	Low	Down
Positive	High	High	Up

Q5 A Explain kernel Trick in support vector machine. [10]

B Explain different ways to combine classifiers. [10]

Q6 Write any **TWO** from the following [20]

A Explain multiclass classification techniques.

B Explain in detail Principal Component Analysis for Dimensionality reduction

C Explain DBSCAN algorithm along with example

****

3 Hours **80 Marks**

1. Question no. 1 compulsory.
2. Answer any three questions out of remaining five
3. Attempt sub questions in order
4. Figures to the right indicate full marks.

1. Write short notes on **(any 4)** **[20]**

a) Capacity building
b) Functions of NIDM
c) Sea walls, embankments and bio shields
d) Triage
e) Environmental hazard
f) National Disaster Management Policy
g) Community based disaster preparedness

2. a) Discuss the framework for disaster management in India. **[8]**

2. b) Explain global warming and climate change. **[6]**

2. c) Comment on radiation hazards. Also discuss possibilities of chemical spills in Mumbai. **[6]**

3. a) Discuss the various types of technological disasters and highlight the specific efforts to mitigate such disasters in India. **[8]**

3. b) Explain the role of various international agencies for Disaster Management. **[6]**

3. c) Explain various means of raising finance for mitigating and managing disasters **[6]**

4. a) Differentiate between structural and non-structural measures of flood mitigation and discuss the importance of forecasting, warning and monitoring system in India **[8]**

4. b) Appraise the role of GIS and GPS in disaster management **[6]**

4. c) Discuss various types of natural disasters in India and highlight their impacts on life. **[6]**

5. a) Explain in detail the design concepts involved in as well as the construction materials used for the safe construction of facilities in case of earthquakes and cyclones. Also discuss the fire resistant facilities that need to be essentially provided in a building/industry. **[8]**

5. b) Elaborate the guidelines laid down by NDMA for disaster management in India. **[6]**

5. c) Explain in detail, vulnerability, with reference to floods and cyclones. List down the preparatory measures for minimizing vulnerabilities related to Tsunami. **[6]**

6. a) Discuss in brief the Disaster Management Act 2005. **[6]**

6. b) Explain Community based disaster preparedness **[5]**

6. c) Is rapid depletion of ground water a type of disaster? To which category does this belongs? What are the reasons for this problem? **[5]**

6. d) Identify and discuss the various hazards which are associated with volcanic eruptions **[4]**

Duration: 3hrs **[Max Marks:80]**

N.B. : (1) Question No 1 is Compulsory.
(2) Attempt any three questions out of the remaining five.
(3) All questions carry equal marks.
(4) Assume suitable data, if required and state it clearly.

1		Attempt any FOUR	[20]
	a	Explain how criminals plan the attack	
	b	Explain various security challenges posed by mobile devices	
	c	Explain need of Cyber law in India	
	d	Explain E-contracts and its different types.	
	e	What are Botnets? How it is exploit by attacker to cause cyber-attack?	
2	a	Explain the classification of cybercrimes with examples.	[10]
	b	Explain Phishing and Identity theft in detail.	[10]
3	a	Explain different buffer overflow attacks also explain how to mitigate buffer overflow attack	[10]
	b	Explain electronic banking in India and what are laws related to electronic banking in India	[10]
4	a	What do you understand by DOS and DDOS attack? Explain in detail.	[10]
	b	Write a note on Intellectual Property Aspects in cyber law.	[10]
5	a	Explain SQL injection attack. State different countermeasure to prevent the attack.	[10]
	b	Explain the objectives and features of IT Act 2000	[10]
6	a	Explain the term evidence and different types of evidences	[10]
	b	Write key IT requirements for SOX and HIPAA.	[10]

****

Duration: - 3 Hours **Marks: 80 Marks**

NB: - Question 1 is compulsory
Solve any four questions from Question no. 1 .
Solve any three questions from the remaining.

1	a.	Define information retrieval and list down classification of information retrieval systems?	**20 (4x5)**
	b.	Explain the process of Structured Text retrieval model.	
	c.	Explain the taxonomy of Information retrieval Model.	
	d.	Explain the role of pattern matching in Information retrieval.	
	e.	Explain multimedia indexing approach ?	
2	a.	Illustrate information retrieval system? Discuss its relationship to DBMS, digital libraries and data warehouses?	**10**
	b.	Explain in detail about vector-space retrieval models with an example?Is the vector space model always superior to the Boolean model?	**10**
3	a.	What is local and global analysis and Differentiate between automatic local analysis and global analysis?	**10**
	b.	What is the role of suffix array and suffix tree in information retrieval system with example.	**10**
4	a.	What is the inverted file? Explain in detail with example.	**10**
	b.	What is the significance tf-idf weight? Can the tf-idf weight of a term in a document exceed 1? Why?	**10**
5	a.	Compare and contrast evaluation of ranked and unranked Retrieval Results ?	**10**
	b.	Define multimedia information retrieval. Discuss indexing and searching?	**10**
6		Write short notes on any two	**20**
	a.	Inexact top K document retrieval	
	b.	Parametric and zone indices	
	c.	Latent Semantic Indexing Model	
	d.	Flat browsing vs hypertext browsing model.	

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

N.B.: 1. Question No. 1 is compulsory.
2. Answer any three out of the remaining questions.
3. Assume suitable data if necessary.
4. Figures to the right indicate full marks.

Q1. Attempt the following (Any 4): **(20)**

a. Explain the concept of UTXO model of Bitcoin.
b. Differentiate between hot and cold wallets
c. Explain mining pool and its difficulty.
d. Compare and contrast private and public blockchain.
e. List and explain various types of nodes used in ethereum.

Q2. Attempt the following:

a. Explain the function of state machine replication. Explain with respect to crowd funding application. **(10)**
b. Compare BFT and PBFT Consensus in detail. **(10)**

Q3. Attempt the following:

a. Compare the role of MSP and Fabric CA. Explain their role in Hyperledger blockchain. **(10)**
b. Explain ethereum architecture and workflow in detail. **(10)**

Q4. Attempt the following:

a. List and explain the types of test networks used in ethereum. **(10)**
b. Explain different visibility specifier of functions in solidity with example. **(10)**

Q5. Attempt the following:

a. What is transaction structure? Explain transaction life cycle in detail. **(10)**
b. Explain the role of address and address payable in solidity with example. **(10)**

Q6. Write short notes on (Any 2): **(20)**

a. Consensus in Bitcoin
b. Hyperledger Fabric
c. Cryptography in Blockchain
d. Defi Architecture

****

(3 Hours) **(TotalMarks:80)**

N.B. : (1) Question No 1 is Compulsory.
(2) Attempt any **three** questions out of the remaining **five**.
(3) All questions carry equal marks.
(4) Assume suitable data, if required and state it clearly.

Q1		Solve any **Four** from the following	[20]
	A	Explain with example Union and Intersection fuzzy operations.	
	B	Determine α-level sets and strong α-level sets for the following fuzzy sets. A = {(B1, 0.4), (B2, 0.3), (B3, 0.9), (B4, 0.1), (B5, 1), (B6, 0.2)} where α = 0.4.	
	C	State PEAS Description for flower pot.	
	D	Design 2-input NAND GATE using McCulloch Pitt's model.	
	E	Explain triangular and T-shape membership functions of fuzzy logic.	
Q2	A	Give the problem formulation for the 8 puzzle problem	[10]
	B	Define Planning and explain partial order planning	[10]
Q3	A	Explain resolution along with example.	[10]
	B	Explain Genetics algorithm along with example.	[10]
Q4	A	Give any five defuzzification methods along with an example	[10]
	B	Explain delta learning rule along with an example.	[10]
Q5	A	Explain the Wumpus World Problem along with all its PEAS properties.	[10]
	B	Explain Vacuum Cleaning machine problem for 4- input and 3-output descriptors using fuzzy logic control systems	[10]
Q6		Write Short notes on following **(Any Two)**	[20]
	A	Simulated Annealing.	
	B	Goal based agent.	
	C	Hill climbing	

****

Duration: 3hrs **[Max Marks: 80]**

N.B.: (1) Question No 1 is Compulsory.
(2) Attempt any three questions out of the remaining five.
(3) All questions carry equal marks.
(4) Assume suitable data, if required and state it clearly.

1 Attempt any FOUR **[20]**

a What is the rule-based and stochastic part of speech taggers?

b Explain Good Turing Discounting?

c Explain statistical approach for machine translation.

d Explain with suitable example the following relationships between word meanings: Hyponymy, Hypernymy, Meronymy, Holynymy

e What is reference resolution?

2 **a** Explain FSA for nouns and verbs. Also Design a Finite State Automata (FSA) for the words of English numbers 1-99. **[10]**

b Discuss the challenges in various stages of natural language processing. **[10]**

3 **a** Consider the following corpus **[10]**

<s> the/DT students/NN pass/V the/DT test/NN<\s>
<s> the/DT students/NN wait/V for/P the/DT result/NN<\s>
<s> teachers/NN test/V students/NN<\s>

Compute the emission and transition probabilities for a bigram HMM. Also decode the following sentence using Viterbi algorithm.
"The students wait for the test"

b What are five types of referring expressions? Explain with the help of example. **[10]**

4 **a** Explain dictionary-based approach (Lesk algorithm) for word sense disambiguation (WSD) with suitable example. **[10]**

b Explain the various challenges in POS tagging. **[10]**

5 **a** Explain Porter Stemming algorithm in detail. **[10]**

b Explain the use of Probabilistic Context Free Grammar (PCFG) in natural language processing with example. **[10]**

6 **a** Explain Question Answering system (QAS) in detail. **[10]**

b Explain how Conditional Random Field (CRF) is used for sequence labeling. **[10]**

****

Time: 03 Hours **Marks: 80**

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

Q1 a) What is the basic difference between traditional RDBMS and Hadoop? [5]

b) What are the 3 V's of big data? Give two big data case studies indicating respective V's with justification. [5]

c) Explain how node failure is handled in Hadoop. [5]

d) List down all six constraints that must be satisfied for representing a stream by buckets using DGIM algorithm with examples. [5]

Q2 a) Describe the four ways by which big data problems are handled by NoSQL. [10]

b) Write a map reduce pseudo code to multiply two matrices. Apply map reduce working to perform following matrix multiplication. [10]

$$M = \begin{matrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{matrix} \quad X \quad V = \begin{matrix} 1 \\ 2 \\ 3 \end{matrix}$$

Q3 a) Suppose the stream is S = {4, 2, 5 ,9, 1, 6, 3, 7}. Let hash functions h(x) = x + 6 mod 32 for some a and b, treat result as a 5-bit binary integer. Show how the Flajolet- Martin algorithm will estimate the number of distinct elements in this stream. [10]

b) i. Create a data frame from the following 4 vectors and demonstrate the output: [10]

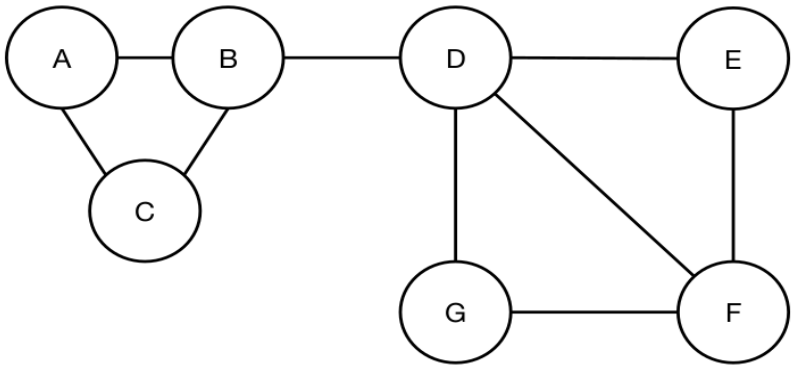
```
emp_id = c (1:5)
emp_name = c("Rick","Dan","Michelle","Ryan","Gary")
start_date = c("2012-01-01", "2013-09-23", "2014-11-15", "2014-05-11", "2015-
        03-27")
salary = c(60000, 45000, 75000, 84000, 20000)
```

ii. Display structure and summary of the above data frame.
iii. Extract the emp_name and salary columns from the above data frame.
iv. Extract the employee details whose salary is less than or equal to 60000.

Q4 a) Explain Map Reduce execution pipeline with suitable example [10]

b) Explain DGIM algorithm for counting ones in a stream with example. [10]

Q5 a) Determine communities for the given social network graph using Girvan-Newman algorithm. [10]



b) List and explain various functions that allow users to handle data in R workspace with appropriate examples. [10]

Q6 a) i. What are the advantages of using functions over scripts? [10]

ii. Suppose you have two datasets A and B.
Dataset A has the following data: 6 7 8 9.
Dataset B has the following data: 1 2 4 5.
Which function is used to combine the data from both datasets into dataset C.
Demonstrate the function with the input values and write the output.

b) How recommendation is done based on properties of the product? Explain with the help of an example. [10]

****