

Duration: 3 Hours

[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.

Q1 Attempt any **FOUR** **[20]**

- A Explain Training error and Generalization error.
- B Differentiate between Supervised and unsupervised Learning
- C Differentiate between Linear regression and Logistic regression.
- D Explain issues in Machine learning.
- E Explain performance evaluation metrics for the classification.

Q2 A Demonstrate MST algorithm along with example. **[10]**

B Explain Logistics regression and performance evaluation metrics. **[10]**

Q3 A Demonstrate steps to design a Machine Learning application. **[10]**

B What is over fitting, under fitting and Bias variance trade-off with reference to Machine learning? **[10]**

Q4 A Demonstrate Ensemble learning based Random Forest algorithm in detail. **[10]**

B Suppose we want Gini index to decide whether the car will be stolen or not. The target classification is "car is stolen?" which can be Yes or No, create a decision tree for the given data below. **[10]**

Car no	Colour	Type	Origin	Stolen ?	Car no	Colour	Type	Origin	Stolen ?
1	Red	Sports	Domestic	Yes	6	Yellow	SUV	Imported	No
2	Red	Sports	Domestic	No	7	Yellow	SUV	Imported	Yes
3	Red	Sports	Domestic	Yes	8	Yellow	SUV	Domestic	No
4	Yellow	Sports	Domestic	No	9	Red	SUV	Imported	No
5	Yellow	Sports	Imported	Yes	10	Red	Sports	Imported	Yes

Q5 A Give steps to design PCA dimensional reduction technique along with an example. **[10]**

B Demonstrate DBSCAN algorithm along with example. **[10]**

Q6 Write detailed note on following. (Any TWO) **[20]**

- A Write a short note on XGBoost ensemble method.
- B Explain support vector machine as constraint optimization problem.
- C SVM Kernel trick

Duration: 3Hours

[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 How Does IT Impact Organizations?
 - b Compare between mobile applications like WhatsApp and Telegram on various parameters like multimedia, security, privacy, storage, data usage, notifications etc.
 - c Describe the tools that augment the traditional SDLC.
 - d Analyze how social computing is inspiring customer service?
 - e Categorize the approach to managing information across an entire organization
- 2 a Analyze the impact of BI on Decision making. [10]**
b Explain Data warehouse in an organization. [10]
- 3 a Identify the measures to improve cyber security with example. [10]**
b Highlight how company can use Big data to gain competitive advantage? [10]
- 4 a Identify the measures to improve cyber security with example. [10]**
b Explain the steps involved in knowledge capturing. [10]
- 5 a List down the types of support provided by Accounting IS, finance IS, production/operations management (POM) IS, marketing IS, and human resources IS. [10]**
b Analyse the main reasons of Computer Crimes. [10]
- 6 a Explain the impact of information system on organization and society. [10]**
b Write note on e-commerce. [10]

Duration: 3 Hours

[Max Marks: 80]

NB:

- (1) Question No.1 is Compulsory
- (2) Attempt any three questions out of 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 are technical measures of Information Security
 - b. Explain concept of Identification, Authentication.
 - c. Illustrate various XSS attacks.
 - d. Explain benefits of Cloud computing
 - e. List down various Malwares and describe anyone.
 2. a. Illustrate an Overview of Certifiable Standards in (How, What, When, Who).term. **[10]**
b. Compare Intrusion Detection System (IDS) and Intrusion Prevention System (IPS). **[10]**
 3. a. Explain how availability is calculated including metrics such as Mean Time Between Failure (MTBF) and Mean Time to Repair (MTTR). **[10]**
b. Describe OCTAVE: Risk Assessment Framework along with its benefits. **[10]**
 4. a. Describe various Disaster Recovery Techniques. **[10]**
b. How are access control methods categorized into administrative, physical, technical and the layering of access control? **[10]**
 5. a. Compare the quantitative and qualitative risk assessment approaches. **[10]**
b. Explain Open Web Application Security Project (OWASP) and identify the common issues in Web Apps. **[10]**
 6. a. What are the key characteristics of NIST risk assessment framework? **[10]**
b. What are the objectives of IT ACT? Explain in detail IT ACT 2000 and IT ACT 2008. **[10]**
-

Time: 3 Hours

Total Marks: 80

N.B. (1) Question No.1 is compulsory

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

- Q.1 (a) What type of data analytics is used in healthcare? (5)
(b) Which imaging technologies do not use radiation? Explain those technologies in brief. (5)
(c) What you mean by the term Natural Language Processing for clinical/medical text data (5)
(d) Define Advanced Data Analytics for Healthcare with 6(six) real-world applications. (5)
- Q.2 (a) How could advances in genome analysis lead to more personalised medicine? (10)
(b) What is visualization? Explain different types of visualization techniques, tools with Advantages and Disadvantages. (10)
- Q.3 (a) Illustrate Predictive Modelling in Healthcare with few examples (atleast 2) (10)
(b) Describe: - (10)
1. BAN
2. Dense/Mesh area network for smart living environment
3. SENSOR TECHNOLOGY
4. Image Registration
5. Feature Extraction
- Q.4 (a) What is validation of clinical prediction models? (10)
(b) Explain types of Fraud detection in healthcare with the help of example. (10)
- Q.5 (a) What are the challenges one may face while processing covid clinical reports? (10)
(b) Define Data science with applications of healthcare data analytics. (10)
- Q.6 (a) Illustrate the process of information extraction from clinical text. (10)
(b) Define Biomedical Imaging Modalities with their Applications. (10)

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 word sense disambiguation?
b Explain reference resolution in detail
c Explain rule-based machine translation systems
d What is hybrid POS tagging?
e Differentiate between Syntactic ambiguity and Lexical Ambiguity
- 2 a Design FST for regular and plural nouns. [10]
b Explain the preprocessing operations in natural language processing [10]
- 3 a Consider the following corpus [10]
<s> a/DT dog/NN chases/V a/DT cat/NN </s>
<s> the/DT dog/NN barks/V loudly/RB </s>
<s> a/DT cat/NN runs/V fast/RB </s>
Compute the emission and transition probabilities for a bigram HMM.
Also, decode the following sentence using the Viterbi algorithm.
The cat chases the dog.
b Compare and contrast Hobbs' Algorithm and Centering Theory. [10]
- 4 a Explain how the supervised learning approach can be applied for word sense [10]
disambiguation
b Explain the N-gram language model and its application. [10]
- 5 a Explain the Porter Stemming algorithm in detail. [10]
b Construct a parse tree for the following sentence using the given CFG rules: [10]
The tall girl sings.
Rules: $S \rightarrow NP VP$
 $NP \rightarrow Det Adj N \mid Det N$
 $VP \rightarrow V \mid V NP$
Det \rightarrow "the"
Adj \rightarrow "tall"
N \rightarrow "girl"
V \rightarrow "sings"
- 6 a Explain text summarization in detail [10]
b Explain how Maximum Entropy is used for sequence labeling. [10]

(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. What is a Merkle tree? Explain the structure of a Merkle tree.
- b. Compare Bitcoin and Ethereum
- c. Differentiate between PoW and PoS.
- d. What is a Smart Contract? What are the different types of smart contracts?
- e. Explain five challenges of Blockchain Implementation.

Q2. Attempt the following:

- a. Explain the components of Blockchain. (10)
- b. What is RAFT consensus algorithm? Explain in detail. (10)

Q3. Attempt the following:

- a. What is cryptocurrency? Explain different types of Cryptocurrencies. (10)
- b. Explain state machine replication. (10)

Q4. Attempt the following:

- a. Explain the following terms with respect to Ethereum: Miner and Mining Node, Gas, Accounts, Ether, Transactions. (10)
- b. Write a program in solidity to implement multi-level inheritance. (10)

Q5. Attempt the following:

- a. What is double spending problem? Which algorithm helps in solving it and how? (10)
- b. Explain view and pure functions in solidity with suitable examples. (10)

Q6. Write short notes on (any 2): (20)

- a. Quorum
- b. Hyperledger Fabric v1 Architecture
- c. PAXOS consensus algorithm
- d. Corda

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 elaborate on the taxonomy of Information retrieval Model. **20**
- b. Explain multimedia indexing approach.
- c. Explain the process of Structured Text retrieval model.
- d. Specify the significance of User Relevance feedback in an IR system.
- e. Specify the need for pattern matching in Information retrieval.
- 2 a. Draw the taxonomy of IR models and explain any one IR modeling technique. **10**
- b. Explain query processing in context of distributed IR **10**
- 3 a. Compare and contrast between automatic local analysis and global analysis? Elaborate on any one with an example **10**
- b. Define what is meant by an Inverted file with an example. Explain it with a neat Diagram. **10**
- 4 a. Elaborate on how the ranked and unranked retrieval sets are evaluated. **10**
- b. Explain the method for calculating tf and idf in a vector model. Further elaborate the need for tf and idf in an information system **10**
- 5 a. With neat diagrams, elaborate on the role of suffix array and suffix tree in any given information retrieval system. **10**
- b. Specify what is meant by Latent Semantic Indexing Model. Further, mention the advantages of Latent Semantic Indexing Model **10**
- 6 Write short notes on **any two** **20**
 - a. Significance tf-idf weight
 - b. Flat browsing vs Hypertext Browsing model.
 - c. Parametric and zone indices
 - d. Roccio method for query expansion

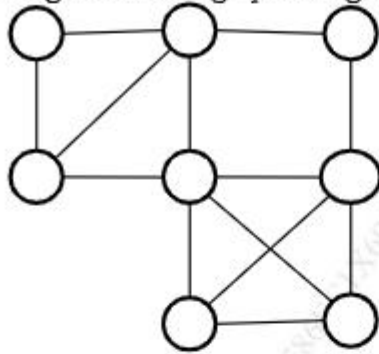
Time: 3 Hours

Total 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.**

- Q.1** a) Explain CAP. How is CAP different from ACID property in databases? [05]
- b) Secondary Name node is a backup of Name node. Is this statement True or False? Justify your answer. [05]
- c) List and explain the core business drivers behind the NoSQL movement. [05]
- d) List down any five constraints that must be satisfied for representing a stream by buckets using DGIM algorithm with examples. [05]
- Q.2** a) List the architectural patterns in NoSQL databases. Discuss the Key-Value and Document-Oriented patterns, focusing on their characteristics, use cases, and examples. [10]
- b) Write a map reduce pseudo code for word count problem. Apply map reduce working on the following document: [10]
- “This is NoSQL. NoSQL handles complex data.”
- Q.3** a) Explain Map Reduce execution pipeline with suitable example. [10]
- b) Create a Bloom filter with the following parameters: [10]
- Size of the bit array $m=8$
Hash functions:
- $$h_1(x) = x \bmod m$$
- $$h_2(x) = (2x+1) \bmod m$$
- $$h_3(x) = (3x+2) \bmod m$$
- (i) Insert the following elements into the Bloom filter: 12, 25, 30, 5
(ii) Check if the following elements are present in the Bloom filter: 6, 55
(iii) Discuss the results of your checks, identifying which elements is true positive and which is true negative.
- Q.4** a) For the stream of integers: 9, 8, 7, 6, 5, 4, 3, 2. Use the hash function, $h(x) = (2x+1) \bmod 32$ and treat the result as a 5-bit binary integer. Show the steps of the Flajolet-Martin algorithm to estimate the number of distinct elements in this stream. [10]
- b) Draw a diagram of the typical Hadoop Ecosystem and explain any two components of it. [10]

Q.5 a) Write an algorithm for the Clique Percolation Method and discover the communities in the given below graph using Clique Percolation Method with clique $k=3$. [10]



b) i. List and explain the functions provided by R to combine different sets of data. [10]
 ii. Write the script to sort the values contained in the following vector in ascending order and descending order: (46, 23, 15, 38, 98, 56, 28, 78). Demonstrate the output.

Q.6 a) The project manager at ABC Corp, Mr. Thomas, needs to track information about ongoing projects in the organization. He has the following details about current projects in a table format: [10]

ProjectId	ProjectName	Budget
1	Website Redesign	150000
2	Mobile App Launch	100000
3	Data Migration	80000
4	AI Development	200000
5	Cybersecurity Audit	50000

i) Create a Data frame in R for the above project data and display the output.
 ii) Show the structure and summary statistics of the Data Frame created.

b) Justify the use of a Content-Based Recommendation System with a specific case study. [10]
