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   [10]
         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.

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   [10]
         validation in detail.

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   [10]
         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.

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

****

309CBE773807C22D23F87371AC5DCE83

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 the challenges faced by Knowledge management in different business scenarios.	
	b	Identify the five factors that contribute to the increasing vulnerability of information resources, and provide a specific example of each one?	
	c	Analyze the impact of BI on Decision making.	
	d	Explain the applications of computer networks.	
	e	Explain the importance of Information systems to Society.	
2	**a**	Develop the plan for delivery application in M-commerce using social computing.	**[10]**
	b	Explain Data warehouse and Data Mart in an organization.	**[10]**
3	**a**	Explain the major security threats to information security and discuss the measures for controlling the same.	**[10]**
	b	Explain CRM. Describe the different types of CRM with example.	**[10]**
4	**a**	What is Cloud Computing? Explain its models.	**[10]**
	b	Write note on e-business.	**[10]**
5	**a**	Identify the measures to improve cyber security with example.	**[10]**
	b	Explain the phases of the system development lifecycle with example.	**[10]**
6	**a**	Explain the steps involved in knowledge capturing.	**[10]**
	b	Compare and contrast Web 1.O,2. O,3. O with example.	**[10]**

****

36858

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]**

****

41703

(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   **(10)**
        crowd funding application.
      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   **(10)**
        blockchain.
      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

****

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.		**10**
		Define multimedia information retrieval. Discuss indexing and searching?	
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.	

_____

3564192C60C540B03A2807A9D29864D3

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   [5]
with justification.

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   [5]
buckets using DGIM algorithm with examples.

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   [10]
working to perform following matrix multiplication.

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

Q3 a) Suppose the stream is S = {4, 2, 5 ,9, 1, 6, 3, 7}. Let hash functions $h(x) = x + 6$ [10]
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.

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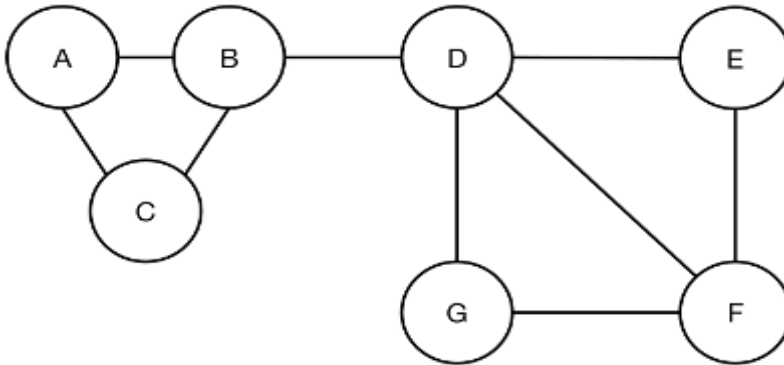
b) List and explain various functions that allow users to handle data in R workspace   [10]
with appropriate examples.

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 [10]
of an example.

****

Time: (3 Hours) [Total Marks: 80]

N.B.: (1) Question **No.1** is **Compulsory**.
  (2) Attempt **any three** questions from the **remaining** questions.
  (3) Assume **suitable** data wherever required but **justify** the same.
  (4) **Figures** to the **right** indicate **full marks**.
  (5) Answer to each new question must start on a **fresh page**.

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 six real-world applications. **[5]**

2. (a) Define Phenotyping Algorithms with key aspects. **[10]**
    (b) What is visualization? Explain different types of visualization techniques, tools with advantages and disadvantages. **[10]**

3. (a) Illustrate Predictive Modelling in Healthcare with at least two examples. **[10]**
    (b) Describe the following: - **[10]**

      1. BAN
      2. Dense/Mesh area network for smart living environment
      3. Senor Technology
      4. Image Registration
      5. Feature Extraction

4. (a) What are the components of EHR? What are the barriers for adopting EHR? **[10]**
    (b) Explain types of Fraud detection in healthcare with the help of example. **[10]**

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]**

6. (a) How will we analyze Mental health status of someone using their tweets on twitter? **[10]**
    (b) Define Biomedical Imaging Modalities with their Applications. **[10]**

3A3FEA51C25E3D7B09A344246F19C135

Duration: 3hrs   **[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. Discuss CIA Triad in Information Security.
    b. Explain concept of High Availability.
    c. Illustrate various XSS attacks
    d. Explain Information Security issues in Cloud computing
    e. Explain various threats to Access Control.

2. a. Describe Risk assessment techniques outlined in ISO31010 framework.   **[10]**
    b. Define Intrusion Detection System. Explain in detail IDS techniques.   **[10]**

3. a. Explain Availability, Mean Time Between Failure (MTBF), Mean Time to Repair
       (MTTR), and Calculate the Availability for a product has MTBF of 200hrs and
       MTTR of 10 hrs.   **[10]**
    b. Explain in detail COBIT Framework.   **[10]**

4. a. Describe various Disaster Recovery Techniques.   **[10]**
    b. Explain any two different Access Control Models from the following.   **[10]**
        a. Discretionary,
        b. Mandatory,
        c. Role based
        d. Rule-based.

5. a. Compare the quantitative and qualitative risk assessment approaches.   **[10]**
    b. Explain various types of Audits in Windows Environment.   **[10]**

6. a. What are the key characteristics of OCTAVE approach?   **[10]**
    b. What are the objectives of IT ACT? Explain in detail IT ACT 2000 and IT
       ACT 2008.   **[10]**

_____