BE / CMPN / SEM-VII / C-2019 / DEC. 2022

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.

Q1. Solve any **four** from following. [20]
    a.   What are the issues in Machine learning?
    b.   Explain Regression line, Scatter plot, Error in prediction and Best fitting line.
    c.   Explain the concept of margin and support vector.
    d.   Explain the distance metrics used in clustering.
    e.   Explain Logistic Regression

Q2. a. Explain the steps of developing Machine Learning applications. [10]
    b. Explain Linear regression along with an example. [10]

Q3. a. Create a decision tree using Gini Index to classify following dataset. [10]

Sr. No.	Income	Age	Own Car
1	Very High	Young	Yes
2	High	Medium	Yes
3	Low	Young	No
4	High	Medium	Yes
5	Very High	Medium	Yes
6	Medium	Young	Yes
7	High	Old	Yes
8	Medium	Medium	No
9	Low	Medium	No
10	Low	Old	No
11	High	Young	Yes
12	Medium	Old	No

    b. Describe Multiclass classification. [10]

Q4. a. Explain the Random Forest algorithm in detail. [10]
    b. Explain the different ways to combine the classifiers. [10]

Q5. a. Compute the Linear Discriminant projection for the following two-dimensional [10]
    dataset. $X1 = (x1, x2) = \{(4,1), (2,4), (2,3), (3,6), (4,4)\}$ and
        $X2 = (x1, x2) = \{(9,10), (6,8), (9,5), (8,7), (10,8)\}$
    b. Explain EM algorithm. [10]

Q6. Write detailed note on following. (**Any two**) [20]
    a. Performance Metrics for Classification
    b. Principal Component Analysis for Dimension Reduction
    c. DBSCAN

****

BE | Sem - VII (CC-2019) | CMPN | Dec-2022

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 function of Map Tasks in the Map Reduce framework? Explain with the help of an example. [5]

b) Demonstrate how business problems have been successfully solved faster, cheaper and more effectively considering NoSQL Google's MapReduce case study. Also illustrate the business drivers and the findings in it. [5]

c) Why is HDFS more suited for applications having large datasets and not when there are small files? Elaborate. [5]

d) Explain the concept of bloom filter with an example [5]

Q2 a) Name the three ways that resources can be shared between computer systems. Name the architecture used in big data solutions and describe it in detail. [10]

b) Write a map reduce pseudo code for word count problem. Apply map reduce working on the following document: [10]

"This is an apple. Apple is red in color".

Q3 a) Suppose the stream is 1, 3, 2, 1, 2, 3, 4, 3, 1, 2, 3, 1. Let $h(x) = 6x + 1 \mod 5$. Show how the Flajolet- Martin algorithm will estimate the number of distinct elements in this stream. [10]

b) Consider the following data frame given below: [10]

subject	class	marks
1	1	56
2	2	75
3	1	48
4	2	69
5	1	84
6	2	53

i. Create a subset of subject less than 4 by using subset () function and demonstrate the output.

ii. Create a subset where the subject column is less than 3 and the class equals to 2 by using [ ] brackets and demonstrate the output.

Q4 a) What are the Core Hadoop components? Explain in detail. [10]

b) With a neat sketch, explain the architecture of the data-stream management system. [10]

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

G-P-code

15066
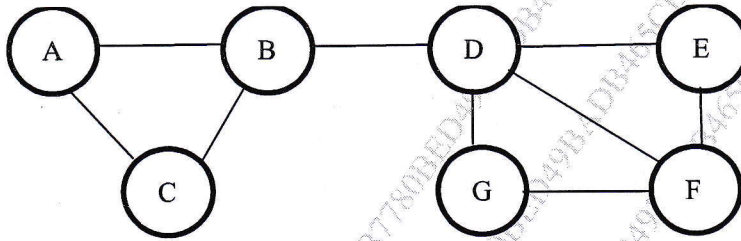
b) The data analyst of Argon technology Mr. John needs to enter the salaries of 10   [10]
employees in R. The salaries of the employees are given in the following table:

Sr. No.	Name of employees	Salaries
1	Vivek	21000
2	Karan	55000
3	James	67000
4	Soham	50000
5	Renu	54000
6	Farah	40000
7	Hetal	30000
8	Mary	70000
9	Ganesh	20000
10	Krish	15000

i.   Which R command will Mr. John use to enter these values demonstrate the output.

ii.   Now Mr. John wants to add the salaries of 5 new employees in the existing table,
which command he will use to join datasets with new values in R. Demonstrate the
output.

Q6 a)   i. Write the script to sort the values contained in the following vector in ascending   [10]
order and descending order: (23, 45, 10, 34, 89, 20, 67, 99). Demonstrate the
output.
ii. Name and explain the operators used to form data subsets in R.

b)   How recommendation is done based on properties of product? Elaborate with a   [10]
suitable example.

BE/CMPN/SEM-VII/C-2019/DEC. 2022

Time: 3 Hours   **Max. Marks: 80**

N.B. (1) Question No. 1 is compulsory
     (2) Assume suitable data if necessary
     (3) Attempt any three questions from remaining questions

Q.1	Any Four	**20[M]**
a	Differentiate between Syntactic ambiguity and Lexical Ambiguity.	**[5M]**
b	Define affixes. Explain the types of affixes.	**[5M]**
c	Describe open class words and closed class words in English with examples.	**[5M]**
d	What is rule base machine translation?	**[5M]**
e	Explain with suitable example following relationships between word meanings. Homonymy, Polysemy, Synonymy, Antonymy	**[5M]**
f	Explain perplexity of any language model.	**[5M]**

Q.2 a) Explain the role of FSA in morphological analysis?

Q.2 b) Explain Different stage involved in NLP process with suitable example. **[10M]**

Q.3 a) Consider the following corpus   **[5M]**
<s> I tell you to sleep and rest </s>
<s> I would like to sleep for an hour </s>
<s> Sleep helps one to relax </s>
List all possible bigrams. Compute conditional probabilities and predict the next ord for the word "to".

Q.3 b) Explain Yarowsky bootstrapping approach of semi supervised learning   **[5M]**

Q.3 c) What is POS tagging? Discuss various challenges faced by POS tagging.   **[10M]**

Q.4 a) What are the limitations of Hidden Markov Model?   **[5M]**

Q.4 b) Explain the different steps in text processing for Information Retrieval   **[5M]**

Q.4 c) Compare top-down and bottom-up approach of parsing with example.   **[10M]**

Q.5 a) What do you mean by word sense disambiguation (WSD)? Discuss dictionary based approach for WSD.   **[10M]**

Q.5 b) Explain Hobbs algorithm for pronoun resolution.   **[10M]**

Q.6 a) Explain Text summarization in detail.   **[10M]**

Q.6 b) Explain Porter Stemming algorithm in detail   **[10M]**

****

P.P. CODE
16298

BE / CMPN / SEM-VII / C-2019 / DEC. 2022

(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. Define blockchain? Compare different types of blockchain.
   b. What is a smart contract? How crowdfunding platforms can be managed using smart contracts?
   c. What is a backup in Practical Byzantine Fault Tolerance (PBFT) algorithm?
   d. What is a Merkle tree? Explain the structure of a Merkle tree.
   e. Write a program in solidity to check whether a number is prime or not.

Q2. **Attempt the following:**
   a. State and explain various challenges that occur while implementing blockchain. (10)
   b. What is a double spending problem? How PoW solves the problem of double spending? (10)

Q3. **Attempt the following:**
   a. Compare Bitcoin and Ethereum. How to calculate Mining difficulty in bitcoin (10)
   b. Explain Hyperledger Fabric v1 architecture. (10)

Q4. **Attempt the following:**
   a. Describe the architecture of Ethereum. (10)
   b. Write a program in solidity to implement multi-level inheritance. (10)

Q5. **Attempt the following:**
   a. Explain PAXOS consensus algorithm for a private blockchain. (10)
   b. Explain fixed and dynamic arrays in solidity with suitable examples. (10)

Q6. **Write short notes on (any 2):** (20)
   a. Corda
   b. UTXO model of Bitcoin
   c. Quorum
   d. Fallback function in Solidity

****

Q. P. CODE
16512

B.E|SEM-VII| CMPN| C-2019| Nov- Dec 2022

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 Differentiate between cybercrime and cyber fraud.
   b Explain various threats associated with cloud computing.
   c Explain methods of password cracking
   d Explain E-contracts and its different types.
   e Explain different attack vectors in cyber security

2 a Explain the classification of cybercrimes with examples. [10]
   b Explain various types of credit card frauds [10]

3 a Explain different buffer overflow attacks also explain how to mitigate buffer [10]
     overflow attack
   b Explain electronic banking in India and what are laws related to electronic [10]
     banking in India

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 the objectives and features of IT Act 2000 [10]
   b What are Botnets? How it is exploit by attacker to cause cyber attack? [10]

6 a Explain SQL injection attack. State different countermeasure to prevent the [10]
     attack.
   b Explain what is Information Security Standard and Explain HIPAA act in detail [10]

****

1317AC8D97C16BA5A75386E9BCA04A88

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 are the different types of MIS?	[05]
b	How is data governance achieved in case of MIS?	[05]
c	Analyse briefly to highlight the difference between Web 2.0 and Web 3.0?	[05]
d	Evaluate the MIS Hierarchy to comment on Decision Support System.	[05]
e	List the main difference between Wireless and Wired Technologies?	[05]
2 a	Give an understanding on types of Control to achieve Security.	[10]
b	What is Mobile Commerce? What are the new challenges that it has introduced in business?	[10]
3 a	What do you mean by CRM? Give its types and relate the role of SC on CRM.	[10]
b	What is Data Mart and Data Warehouses? Give two examples which show generation of Big Data.	[10]
4 a	Write short notes on (1) TPS (2) ERP	[10]
b	Evaluate the role of Confidentiality, Integrity and Availability in order to achieve security.	[10]
5 a	What is the need of Social Computing for Businesses?	[10]
b	Create MIS system for any hospital.	[10]
6 a	What is Big Data? What are the various challenges and characteristics of Big Data?	[10]
b	Describe various Cloud Computing Models and highlight their evolution.	[10]

****

1317AC8D97C16BA5A75386E9BCB6AF71