TE-/IT/ 8 Sem VI / R-19- 8-Scheme / May-June 2023.

(3 hours) **[80 marks]**

NOTE:

1. Question No 1 is compulsory
2. Attempt any three questions from remaining.
3. Assume suitable data if necessary and state the same.

Q.1 [20]

A) Draw Data warehousing Architecture?
B) What is noisy data? How to handle noisy data?
C) Compare and contrast between OLTP and OLAP.
D) Explain concept of information gain and gini value used in decision tree algorithm.

Q.2

A) What is Data mining? Explain KDD process with diagram. [10]
B) Consider we have age of 29 participants in a survey given to us in sorted order. [10]
5, 10, 13, 15, 16, 16, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70, 85
Explain how to calculate mean, median, standard deviation, 1st and 3rd Quartile for given data and also compute the same. Show the Box and Whisker plot for this data.

Q.3

A) Explain market Basket Analysis with example. [10]

B) Consider Training dataset as given below. Use Naive Bayes Algorithm to determine whether it is advisable to play tennis on a day with hot temperature, rainy outlook, high humidity and no wind? [10]

Outlook	temperature	Humidity	Windy	Class
sunny	hot	high	false	No
sunny	hot	high	true	No
overcast	hot	high	false	Play
rain	mild	high	false	Play
rain	cool	normal	false	Play
rain	cool	normal	true	No
overcast	cool	normal	true	Play
sunny	mild	high	false	No
sunny	cool	normal	false	Play
rain	mild	normal	false	Play
sunny	mild	normal	true	Play
overcast	mild	high	true	Play
overcast	hot	normal	false	Play
rain	mild	high	true	No

Q.4

A) What is an outlier? Explain various methods for performing outlier analysis. [10]

B) Use the Apriori algorithm to identify the frequent item-sets in the following database. Then extract the strong association rules from these sets.Assume Min. Support = 50% Min. Confidence=75% [10]

Tid	a	b	c	d	e	f	g
Items	1,2,4,5,6	2,3,5	1,2,4,5	1,2,4,5	1,2,3,4,5,6	2,3,4	1,2,4,5

Q.5

A) Cluster the following eight points (with (x, y) representing locations) into three clusters: [10]
A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)
Assume Initial cluster centers are at: A1(2, 10), A4(5, 8) and A7(1, 2).
The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as- $P(a, b) = |x2 - x1| + |y2 - y1|$
Use K-Means Algorithm to find the three cluster centres after the second iteration.

B) Compare star schema, Snow flakes schema and star constellation [10]

Q.6 Write short note on following (Any 4) [20]

A) Dimensional Modeling.
B) Random Forest Technique.
C) Decision Tree Induction.
D) Cross Validation.
E) DBSCAN Algorithm

****

TE-/IT/8 Sem VI / R-19- 8-Scheme / May-June 2023.

(3 hours) **[80 marks]**

NOTE:

1. Question No 1 is compulsory
2. Attempt any three questions from remaining.
3. Assume suitable data if necessary and state the same.

Q.1 [20]

A) Draw Data warehousing Architecture?
B) What is noisy data? How to handle noisy data?
C) Compare and contrast between OLTP and OLAP.
D) Explain concept of information gain and gini value used in decision tree algorithm.

Q.2

A) What is Data mining? Explain KDD process with diagram. [10]
B) Consider we have age of 29 participants in a survey given to us in sorted order. [10]
5, 10, 13, 15, 16, 16, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70, 85
Explain how to calculate mean, median, standard deviation, 1st and 3rd Quartile for given data and also compute the same. Show the Box and Whisker plot for this data.

Q.3

A) Explain market Basket Analysis with example. [10]

B) Consider Training dataset as given below. Use Naive Bayes Algorithm to determine whether it is advisable to play tennis on a day with hot temperature, rainy outlook, high humidity and no wind? [10]

Outlook	temperature	Humidity	Windy	Class
sunny	hot	high	false	No
sunny	hot	high	true	No
overcast	hot	high	false	Play
rain	mild	high	false	Play
rain	cool	normal	false	Play
rain	cool	normal	true	No
overcast	cool	normal	true	Play
sunny	mild	high	false	No
sunny	cool	normal	false	Play
rain	mild	normal	false	Play
sunny	mild	normal	true	Play
overcast	mild	high	true	Play
overcast	hot	normal	false	Play
rain	mild	high	true	No

Q.4

A) What is an outlier? Explain various methods for performing outlier analysis. [10]

B) Use the Apriori algorithm to identify the frequent item-sets in the following database. Then extract the strong association rules from these sets. Assume Min. Support = 50% Min. Confidence=75% [10]

Tid	a	b	c	d	e	f	g
Items	1,2,4,5,6	2,3,5	1,2,4,5	1,2,4,5	1,2,3,4,5,6	2,3,4	1,2,4,5

Q.5

A) Cluster the following eight points (with (x, y) representing locations) into three clusters: [10]
A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)
Assume Initial cluster centers are at: A1(2, 10), A4(5, 8) and A7(1, 2).
The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as- $P(a, b) = |x2 - x1| + |y2 - y1|$
Use K-Means Algorithm to find the three cluster centres after the second iteration.

B) Compare star schema, Snow flakes schema and star constellation [10]

Q.6 Write short note on following (Any 4) [20]

A) Dimensional Modeling.
B) Random Forest Technique.
C) Decision Tree Induction.
D) Cross Validation.
E) DBSCAN Algorithm

****

TE - IT | Sem VI | R-scheme | May - June 2023

(3 hrs.) Maximum Marks = 80

NB:

1. Question No. 1 is compulsory and solve any THREE questions from remaining questions
2. Assume suitable data if necessary
3. Draw clean and neat diagrams

			Marks
Q1.		Attempt any four	
	a.	Explain Semantic Web Stack	5
	b.	Explain an arrow function in TypeScript	5
	c.	What is Routing in AngularJS	5
	d.	Discuss default database in MongoDB : local, admin, and config	5
	e.	Discuss technologies AJAX works with to create interactive web pages	5
Q2.	a.	Define Clickstream Analysis and state its applications.	10
	b.	Explain Modules in TypeScript with example	10
Q3.	a.	Explain AngularJS ng-app, ng-init, ng-model directive with examples	10
	b	Explain Accessing and Manipulating Databases commands in MongoDB	10
Q4	a.	Explain with example concept of Flask Templates	10
	b.	With diagram explain working of AJAX	10
Q5	a.	What are the expressions in AngularJS?	10
	b.	Explain the concept of Collections and Documents in MongoDB with examples	10
Q6	a.	What is Flask URL Building?	10
	b.	Short note on Angular JS Data Binding	10

TE - IT | Sem VI | C- Scheme | 16/05/2023.

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 Forward and Backward chaining
b Compare different search techniques based on their time complexities.
c What is a histogram ? Can we perform univariate graphical analysis using histogram?
d Explain various measures of the central tendencies of a statistical distribution.
e State PEAS of automated taxi driver.
f What are the different ways of knowledge representation?

2 a Can 1liter water be measured using 10 liter and 4 liter jug? Justify. [10]
b Compare Linear Regression Vs Logistics Regression with suitable diagrams and formulas. [10]

3 a State A* algorithm and explain with example how A* searching algorithm helps in finding the goal with optimal path. [10]
b With respect to Quantitative data analysis explain following: [10]
i. Measure of central tendencies
ii. Measure of spread
iii. Skewness and Kurtosis

4 a 1. Marcus was a man. [10]
2. Marcus was a Pompeian.
3. All Pompeians were Romans.
4. Caesar was a ruler.
5. All Pompeians were either loyal to Caesar or hated him.
6. Every one is loyal to someone.
7. People only try to assassinate rulers they are not loyal to.
8. Marcus tried to assassinate Caesar.

Was Marcus loyal to Casear ? Solve using resolution.
b In detail, explain steps in the Data Science Project. [10]

5 a What are the different types of Machine Learning algorithms? Give example of each category. [10]
b Can min-max be used for team games? Draw sample trees for 2 and 3 teams. [10]

75D8E3CF77ED0252F6C15C1FD6EF01BE

6 a Consider you are performing ML for predicting housing prices you have trained three models and following data summarizes the predicted house price by each model for 5 different trial runs. [10]

Model Code	House Price Predicted (Lakh Rs)				
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
A	3.5	3.4	3.8	3.5	3.4
B	3.9	3.8	3.7	3.9	3.6
C	3.5	3.3	3.6	3.5	3.8

Perform One way ANOVA F Test on this data and comment on whether the mean house price predicted by models A, B, C are same with level of significance 0.05. (Use of F Table is allowed)

b What are the rules of conversion from predicate to CNF? Explain each rule with proper example. [10]

****

TE - IT | Sem VI | C - Scheme | 18/05/2023

(3 Hours)

Total Marks: 80

N.B.: (1) Question No.1 is compulsory.
(2) Attempt any three questions from the remaining five questions.
(3) Make suitable assumptions wherever necessary but justify your assumptions.

1.	(a) Explain Mobile forensic. What are various challenges in mobile forensics	05
	(b) Explain Forensic Duplicates as Admissible Evidence.	05
	(c) What is evidence handling procedure?	05
	(d) What are Challenges in network forensics ?	05
2.	(a) Explain Incident Response Process and its methodology.	10
	(b) Compare active attacks vs Passive attacks. Classify the cybercrimes and explain any one briefly.	10
3.	(a) Discuss basic security precautions to be taken to safeguard Laptops and wireless devices and What are the devices related to security issues?	10
	(b) Explain Volatile Data Collection from Windows system	10
4.	(a) What do you understand by social engineering? Give classification	10
	(b) Briefly explain Types of digital Evidence with examples	10
5.	(a) Explain process for collecting Network Based Evidence.	10
	(b) Explain various guidelines for digital forensic report writing along with its goals.	10
6.	Write a short note on (Any Two) (1) Tools used in network forensics (2) Roles of CSIRT in handling incident (3) Email Tracing- Internet Fraud	20

TE- IT | Sem VI | C- Scheme | May-June 2023

Time: 3 HRS **Marks: 80**

Note: 1. Q. No1 is compulsory

2. Solve any three questions out of the remaining five

3. Figures to right indicate full marks

4. Assume suitable data where necessary

Q.No1 Solve any four (20)

a. Explain any four features of MANET and compare MANET and WSN

b. Write note on LTE frame structure in detail

c. Describe evolution of 1G to 5 G mobile systems.

d. Outline the method that supports mobility in CISCO Unified Wireless Network

e. Write note on CDMA2000

2

a. Draw and explain UMTS network architecture and compare GSM and UMTS (10)

b. Draw neat diagram of GPRS system architecture and explain the function of each block in it. (10)

3. a. Give the significance of WEP protocols. What are the features of WPA2 (10)

b. Draw and explain the architecture of Cisco UWN with its features. (10)

4. a. Explain the layered architecture of WSN protocol and discuss issues and challenges in WSN. (10)

b. Draw and explain LoRaWAN network architecture and technology stack in detail. (10)

5. a. What is Zigbee. Explain ZigBee protocol stack in detail? (10)

b. Explain 4G network architecture with its specifications. (10)

6. a. Describe Bluetooth architecture and protocol Stack. Also discuss its limitations. (10)

b. Draw and explain protocol architecture of IEEE802.11. Also discuss on power management in IEEE802.11 infrastructure network. (10)

****

TE IT | Sem VI | H/M | C-Scheme | 22/05/23.

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 Categorical data and quantitative data.

b Find S.D of the average temperature recorded over a five-day period last winter 18,22,19,25,12

c Define Binomial distribution and Poisson distribution.

d Explain Type1 and Type 2 error in detail.

e Define the following key terms for simple linear regression.
i)Response ii) Record iii) Independent variable iv) Regression co-efficient v) Residuals

2 a The runs scored in a cricket match by 11 players are as follows:7,16,121,51,101,81,1,16,9,11,16. Find mean, mode, median for the given data. [10]

b An agent sells life insurance policies to five equally aged healthy people. According to recent data, the probability of a person living in these conditions for 30 years or more is 2/3.Caluclate the probability that after 30 years if [10]
i) All five people are still living.
ii) At least three people are still living.
iii) Exactly two people are still living (Hint: Binomial Distribution)

3 a X is a normally distributed variable with mean $\mu=30$ S. D $\sigma=4$.Find i) $P(X<40)$ ii)$P(X>21)$ iii) $P(30<X<35)$ [10]

b Brief the steps in multinomial distribution goodness of fit. Elaborate the steps with an example. [10]

4 a Brief the steps in test of independence. Elaborate the steps with an example [10]

b Find the simple linear regression that fits the given data and co efficient of determination. [10]

Bill	34	108	64	88	99	54
Tip	5	17	11	8	14	5

5 a In the context of multiple linear regression. Explain what is over fitting and multi collinearity. [10]

b Predict equation for y. [10]

y	x1	x2
-3.7	3	8
3.5	4	5
2.5	5	7
11.5	6	3
5.7	2	1

6 a Explain TIME SERIES PATTERNS [10]

i)Horizontal Pattern ii) Trend Pattern iii)Seasonal Pattern

iv)Trend and Seasonal Pattern v)Cyclical Pattern

b Consider the following time series data. [10]

Week	1	2	3	4	5	6
Value	18	13	16	11	17	14

Using the naive method (most recent value) as the forecast for the next week, compute the following measures of forecast accuracy.

i) Mean absolute error. ii) Mean squared error.

iii) Mean absolute percentage error. iv) Determine the forecast for week 7?

****