Reg. No.					6			
								- 1

B.Tech. DEGREE EXAMINATION, DECEMBER 2023

Sixth Semester

18CSE366J - DATA MINING AND ANALYTICS

(For the candidates admitted from the academic year 2020-2021 & 2021-2022)

Note:
(i)

- **Part A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40th minute.
- (ii) Part B & Part C should be answered in answer booklet.

Time: 3 hours	Max. M	larks	: 100	
$PART - A (20 \times 1 = 20 Marks)$	Mark	s BL	СО	PO
Answer ALL Questions				
1 is a statistical methodology that is used for numer	ic prediction 1	1	1 F(1
(A) Regression analysis (B) Covariance				
(C) Correlation (D) Standard dev	riation			
2. Identify the kind of learning algorithm used for 'Facial id	lentities with facial 1	1	1	1
expressions'.				
(A) Prediction (B) Recognition	patterns			
(C) Recognizing anomalies (D) Generating p	attern		~	
3. OLAP stand for	1	1	1	1
(A) Online analytical processing (B) Online applic	cation processing			
(C) Online application programming (D) Online analy				
4 is not a technique for handling noisy data.	1	1	1	1
	•	•	-	•
(A) Regression (B) Binning (C) KDD process (D) Clustering				
(C) KDD process (D) Clustering				
5. Data visualization techniques are used to	1	1	1	1
(A) Calculate accuracy (B) Detect outlier	rs			
(C) Improve training accuracy (D) Integrate data	ì			
6 obtains a reduced representation of the dataset.	1	1	2	1
	mation			
(A) Data Cleaning(B) Data transfor(C) Data reduction(D) Data integrat	ion			
7. To remove noise and inconsistent data is needed	1	1	2	1
(A) Data cleaning (B) Data transfor	mation			
(C) Data reduction (D) Data integral:	ion			
(b) Data Integral.	ion			
8. The values of a attribute are symbols or names of	f things.	1	2	1
(A) Ordinal (B) Nominal				
(C) Ratio (D) Interval				

9.	bins. is a top	down splitting techn	ique	based on a specified number or	1	1	2	1
	(A) Normalizati	on	(B)	Binning				
-	(C) Clustering	On	` '	Classification				
10.	is the no	ocess of finding a m	odel	that describes and distinguishes	1	1	2	1
	data classes.							
	(A) Association	rule mining	(B)	Data discretization				
	(C) Clustering	C		Classification				
11.		occurs frequency it is			1	1	3	1
				Semi structured pattern				
	(C) Frequent pa	ttern	(D)	Structured pattern				
12.	The apriori algor	ithm is used for	dat	ta miming task.	1	1	3	1
12.	(A) Association			Clustering				
	(C) Classification			Regression				
13.		a supervised learning.			1	1	3	1
	(A) PCA		` '	Naive Bayesian				
	(C) Linear regre	ession	(D)	Decision tree				
14	Decision nodes a	re represented as			1	1	3	1
17.	(A) Disks	to represented us	(B)	Square				
	(C) Circles		(D)	Triangles				
	(0)		(-)					
15.	Which of the foll	owing is a testable hy	pothe	esis?	1	1	3	1
	(A) Dogs are be	etter pets than cats	(B)	Eating vegetables every day				
				improves health				
	(C) The moon i	s made of cheese	(D)	People prefer blue cars over read cars				
16.			widel	y known smoothing model for	1	1	4	1
	forecasting data t		(T)	v* 4.				
	(A) Moving ave	erage		Holt				
	(C) Winter		(D)	Exponential				
17.	models	use past values of the	time	series to predict future values.	1	1	4	1
	(A) Auto regres			Moving average				
	(C) Auto repeti		(D)	Partial auto correlation function				
	. ,		` '					
18.				and its lagged values.	1	1	5	2
	(A) Auto regres			Auto correlation				
	(C) Auto regula	rization	(D)	Regularization				
19	In the regression	equation, $Y = 24 - 3x$	the s	lope is	1	1	5	2
471	(A) 24			 				
	(C) 3			-3				
•	•		` /		1	1	5	1
20.				current data to make predictions.	1	1	J	1
	(A) Data mining(C) Predictive	g		Data pattern				
	(C) Fredictive	•	(U)	Descriptive				

Page 2 of 3

04DA6-18CSE366J

	PART – B ($5 \times 4 = 20$ Marks) Answer ANY FIVE Questions	Marks	BL	CO	PO
21.	What are the major issues in data mining?	4	2	1	2
22.	Differentiate supervised learning and unsupervised learning.	4	2	2	2
23.	How data is transformed in data pre processing method?	4	2	2	1
24.	Write steps involved in KNN algorithm.	4	1	3	1
25.	How logistic regression is different form linear regression?	4	1	5	2
26.	Annotate the different tests of hypotheses.	4	2	4.	2
27.	Write note on moving average.	4	2	5	1
	PART – C ($5 \times 12 = 60$ Marks) Answer ALL Questions	Marks	BL	CO	PO
28. a.	Define machine learning. Discuss different types of algorithm in detail.	12	1	1	2
b.	(OR) Explain various stages of KDD process in detail with neat diagram.	12	2	1	3
29. a.	How can frequent patterns are classified? Explain.	12	2	3	1
	(OR)				
b.	Explain statistical Bayesian classification with real time data set.	12	3	3	2
30. a.	Describe data mining knowledge representation with neat diagram.	12	3	2	1
b.	(OR) Explain about need for data preprocessing and data quality in data mining operations.	12	4	2	1
31. a.	Elucidate different test used to validate logistic regression.	12	2	4	1
b.	(OR) Describe semi parametric and non parametric regression models.	12	2	4	1
32. a.	Enumerate auto covariance and auto correlation in detail.	12	2	5	1
b.	(OR) Discuss ARMA and ARIMA in detail.	12	2	5	1

* * * * *

. .