n No		
Reg. No.		

B.Tech. DEGREE EXAMINATION, MAY 2022

Sixth Semester

		ITAIT	NG AND ANALYTICS ademic year 2018-2019 to 2019-2020,	,			
	18CSE366J - DATA N	TINI	NG AND ANALYTICS ademic year 2018-2019 to 2019-2020, the Good AD minutes and OMR sheet				50 BT
	(For the candidates admitted from the	ic tice	OMR sheet	shoul	d be	hand	ded
Note:	omR she	et wi	thin first 40 minutes				
(i)	Part - A should be answered in Old mit over to hall invigilator at the end of 40th mit over the end of 40th mit	nute.					
	over to hall invigilator at the end of Part - B should be answered in answer boo	klet.					
(ii)	Part - B should be all sweet			Max.	Ma	rks:	75
Time: 21	4 Hours			Marks	BL	со	PO
	$PART - A (25 \times 1 = 2)$	25 M	(arks)				
				1	1	1	1
	Which of the following is an essent	ial p	process in which the intelligent				
1.	Which of the following is an essent	rns?					
	methods are applied to exact data patte	B)	Data mining				
	(A) Watchousing	(D)	Data selection				
	(C) Text mining	()				-	1
	is a repository of multiple he	etero	geneous data sources organized	1	1	1	
2.	under a unified schema at a single si	to to	facilitate management decision				
		10 10	Tacintate manage				
	making.	(D)	Data mining				
	(A) Data wateriouse		KDD				
	(C) DBMS	(D)	KDD				
3	is a dataset which contain	samı	ples that to not share with the	1	2	1	1
J.	common characteristics or model of th	e da	taset.				
	(A) Outliers	(B)	Data discrimination				
	(2.2)		Bounding box				
				- 10			
4.	Which of the following is not an opera	ation	of OLAP?	1	1	1	1
	(A) Drill up	(B)	Roll up				
	(C) Flip up	(D)	Pivot				100
5.	is not a data mining function.						
٥.	(A) Classification	(D)	Calastian - 11 .				
	(C) Characterization and		Selection and interpretation				
	discrimination	(D)	Clustering				
	discrimination						
6.	studies the collection, analy	rsis	interpretation or explanation an				
	presentation of data.	u.u.,	interpretation of explanation an	id !		1	2
	(A) Statistics	(D)	Views P				
	(C) Data mining		Visualization				
	(c) Data mining	(D)	Clustering				
7.	When performing PCA we want to						
	(A) Find orthogonal vectors	(D)	Date		1	1	2
		(B)	dimension the number	of			
	(C) Find the most meaningful	(D)	Find the components of t				
	basis	10000	dataset components of t	he			
Page 1 of 4							

8. The initial steps concerned in the	of knowledge discovery is	1	11 3			
a The initial steps concerned in the	(P) Data integration		3	. 4		19 is a Great corder fermalism and if a
8. The initial steps control	(D) Data transformation					18 is a first order iterative optimization algorithm for finding a local
	(D) Data danstorman					minimum of a differential function.
(C) Data cleaning		- 1				(A) Steepest descent (B) Stochastic descent
9. It is the main technique employed	for data selection	-1-	1 :	2 95		(C) Mini descent (D) Batch descent
o It is the main technique employee	(B) Clustering			. 0		
(A) Noise	(D) Sampling					19 is a forecasting method where historical evidence is unavailable 2 4
(A) Noise	(D) Samples					(A) Quantitative method (B) Average method
(C) Histogram		- 13	19			
partitions the objects into	different groups.	1	1 -	2	17	(C) Qualitative method (D) Naive method
O. partitions the objects	(B) Clustering					
(A) Mapping	(D) Prediction					20 is a statistical measure used to know how data is divided across a 1 4 1
(C) Classification						range.
(C) Classific	in the case of Bayesian network'	7 1	-1	4		(A) Central tendency (B) Measures of variability
at following properties	is false in the case of Bayesian network			4/	1	(C) Measures of frequence (D) Measures of dispersion
Which of the following to						(c) includes (b) includes
						The state of the s
(A) The edges are (B) Contains cycle (C) Represents conditional relation	as among random variables					21. The function is defined as the sequence of covariances of a 1 1 3 1
(B) Possessents conditional relation	IIS among re-					stationary process.
(C) Represent						(A) Auto correlation (B) Auto covariance
(D) Contains edges			140			(C) Partial auto correlation (D) Partial auto covariance
A collection of one or more items	is called as	1	4	2	4	
A collection of one or more items	(B) Support					22 smoothing technique is a widely known smoothing model for
A Concess	(D) Support count					22 smoothing technique is a widely known smoothing model for
(A) Itemset						forecasting data that has a trend.
(C) Confidence	ater than or equal to minimum suppo	0000 100	-			(A) Moving average (B) Holt
	ater than or equal to minimum suppo	ort 1	2	2	1	(C) Winter (D) Exponential
A itemset whose support is gre	arer trian or or					
An itemsee						1 1 3 1
threshold is	(B) Frequent itemset					23 ARIMA stands for
(A) Itemset	(D) Threshold values					
(C) Intrequent items	(D)					moving average average
		- 1	94	3	1	(C) Auto reactive integrated (D) Auto reactive integrated mean
What does FP growth algorithm do	?	and the same				
What does FP growin argentant	as (B) It mines all frequent patter	ns				moving average
(A) It mines all frequent patter	ns (B) It mines all frequent patter th through pruning rules w	ith				24 The living modes are represented as in decision tree.
showing printing inter						
mrough prairie	higher support It mines all frequent patterns constructing an itemset	har				(A) Disks (B) Squares
lesser support	(D) It mines all frequent patterns	by				(C) Circles (D) Triangles
(C) It mines all frequent pattern	constructing an itemset					(C) Circles
						Or disamplifying 1 1 5 3
by constructing a .	direct application of frequent item	100	1	1	1	25. optimization is said to be conflicting objectives.
W 14	times application of frequent item	set				(A) Stochastic (B) Multi-objective
Which of the following is the	direct application					
Which of the	4 4					(C) Non-convex
mining/	(B) Market basket analysis					
(A) Social network analysis	(D) Intrusion detection					Marks BL CO FO
(C) Outliner detection	(D) Intrusion detection					PART – B (5 × 10 = 50 Marks)
C) Outimer detection		1	1	1 0	6 1	Answer ALL Questions
127	a deslana is					Answer ALL Questions
In the regression equation, $y = 24$	3x the stope is					10 3 1 1
III tile regression ad	(B) -24					Sorts machine learning. Discuss various learning algorithms.
(A) 24	(D) -					26. a. State machine learning. Discuss various learning algorithms.
(C) 3	(D) -3				10 V	
()		1		2	4 1	(OR) 10 3 1
		market St.				b. Elaborate the applications of data mining in different fields.
n binary logistic regression	. dable con	sist				b. Elaborate the applications of data and a second
in omary logistic regionals	s (B) The dependent variable con	D, W				Availan 10
A) The dependent variable	of two categories					Discuss the various process used in data reduction.
continuous	of two categories	is				27. a. Discus the various process used in data reduction.
	one on dependent variable	10000				
and the second and second about	(1)) The dependent	100				
continuous C) There is no dependent variable	divided into two equal	sub				(OR)

The state of the s	10 4
b. Illustrate different types of visualization techniques with neat diagram.	10 3 3
b. Illustrate different types of visualization. 28. a. Elaborate how classification is different from clustering. (OR)	10 3
b. Discuss different data mining algorithms in detail.	10 4
b. Discuss different data into a constant of the constant of t	10 4
b. Explain non parametric regression method in detail.	10 3
b. Explain non parametric 122 30. a. Define auto-correlation and explain its properties.	10 4
(OR) b. Explain decision and risk analysis in detail.	
****	1000

Page 4 of 4