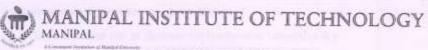
Reg. No.	1
1.00	J



III SEMESTER M.C.A

END SEMESTER EXAMINATIONS,

NOV/DEC 2017

SUBJECT: DATA WAREHOUSING AND DATA MINING (MCA-5102)

REVISED CREDIT SYSTEM (17/11/2017)

Time: 3 Hours

MAX. MARKS: 50

Instructions to Candidates:

- Answer ALL FIVE FULL questions.
- Missing data may be suitable assumed.

1A.	Define Data Mining. Explain with a neat diagram, the steps involved in the Knowledge Discovery in Databases process.	5
1B.	Consider a data set with the values 230,375, 428, 705, 1500. Perform Data transformation on each of the above values with:	3
	(i) the min-max normalization method by setting min = 0 and max = 5	
	(ii) the decimal scaling method	
1C.	What is an ordinal attribute? Give appropriate examples.	2
2A.	The following are the set of 6 transactions representing the purchase of books. Let minimum support be = 2. (i) Find frequent itemsets using FP-growth algorithm or the Apriori algorithm.	5
	(ii) Generate all the association rules with a minimum confidence of 80 %.	

T1	ANN, CC, TC, CG
T2	CC, D, CG
Т3	ANN, CC, TC, CG
T4	ANN, CC, D, CG
T5	ANN, CC, D, TC, CG
T6	CC,D,TC

2B.	What are	the key	features	of a	Data	Warehouse?
-----	----------	---------	----------	------	------	------------

3

2C. What is the need for concept hierarchies? Create a concept hierarchy for the attribute "Order Date".

3A. Details of cars such as, its color, type of car and its origin is recorded. The variable 5 'stolen' indicates whether the car was stolen or not. Predict the class label "Stolen" for a car with attributes – color – Red, type – Sports and origin – Imported using the Naive Bayesian method.

Color	Туре	Origin	Stolen
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Yellow	Sports	Domestic	No
Yellow	Sports	Imported	Yes
Yellow	SUV	Imported	No
Yellow	SUV	Imported	Yes
Yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes

3B.	Ho	w is ribute:	the a	ttribut	te se	lection	mea	sure -	- *Info	rmation	Gai	n" com	puted	for n	umeric	3
3C.													2			
4A.		(i) D	raw a	scatte	er plo		strate	the re	lations	ship beto		SERVICE SERVICE	ilue fo	or x = 1	4.6.	5
	x	3	6	9	8	10	11	12	13	13.5	14	14.5	15	15.2	15.3	
	У	4	5	7	6	8	10	12	14	16	18	22	28	35	42	
4B.	Wh	at st	rategie	es co	ould I	be ad	opted	for s	epara	tion of	test	and tra	aining	datas	set for	3
4B.	1	class	fiers?							tion of) data:	set for	3
10000	List	class t any t nsider	two me	easur	res tha	at qua data po	ntify th	e accu	uracy (ction a	algorithi	ms.		set for	
4C.	List Cor A1	t any i	two me the fo	easur ollowii (2,5)	ng 8 d	at qua data po 8,4), B	ntify th pints w 31 (5,8 he k-n	ie acci vith (x,), B2 (neans	y) repr 7,5), E	of predic	ction a g loca C1 (*	algorithm ation. 1,2) & C	ms. C2 (4,	9)		2
4C.	List Cor A1	t any t nsider (2,10) (i) Ci	two me the fo	easur (2,5) the d C1 to	ng 8 d , A3 (lata u be in	at quadata po 8,4), Busing t	ntify th points w 31 (5,8 he k-n uster c	ith (x,), B2 (y) repr 7,5), E algori	of prediction of	ction a g loca C1 (*	algorithm ation. 1,2) & C	ms. C2 (4,	9)		2
4C.	List Cor A1	class t any t nsider (2,10) (i) Cl	two me the fo	easur ollowin (2,5) the d C1 to e the	ng 8 c , A3 (lata u be in	at qual data po 8,4), B using t itial clu cluster	ntify the points was 1 (5,8) the k-nuster cours (last	vith (x,), B2 (neans centres	uracy (y) repr 7,5), E algori con onl	of prediction of	g loca g loca C1 (1	algorithm ition. 1,2) & C clusters	ms. 22 (4,	9) Iming /		2

MCA-5102