

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY
BE – SEMESTER-VII (NEW SYLLABUS) EXAMINATION- SUMMER - 2018

Subject Code: 2170715

Date: 08/05/2018

Subject Name: Data Mining and Business Intelligence (Department Elective-II)

Time: 02:30 PM TO 05:00 PM

Total Marks: 70

Instructions:

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

		MARKS
Q.1	(a) Explain KDD process using figure.	03
	(b) Do feature wise comparison between BI and DW.	04
	(c) Explain research issues in Data Mining.	07
Q.2	(a) Explain schemas: Stars, snowflakes and fact constellations using figures.	03
	(b) Do feature wise comparison between ROLAP and MOLAP.	04
	(c) Enlist the preprocessing steps with example. Explain procedure of any technique of preprocessing.	07
OR		
	(c) Explain what is concept description? Explain data generalization, summarization-based characterization using example.	07
Q.3	(a) Do feature wise comparison between classification and prediction.	03
	(b) Write a note on incremental Association Rule Mining.	04
	(c) Generate frequent itemsets and generate association rules based on it using apriori algorithm. Minimum support is 50% and minimum confidence is 70%	07

TID	Items
100	1, 3, 4
200	2, 3, 5
300	1, 2, 3, 5
400	2, 5

OR

- | | | |
|------------|--|-----------|
| Q.3 | (a) Differentiate between Overfitting and Tree Pruning w.r.to following parameters.
i). definition figure
ii). use in particular situation | 03 |
|------------|--|-----------|

iii). limitation

- (b) Explain Mining Multiple-Level Association Rules using example. **04**
- (c) Generate decision tree using CART algorithm for the following dataset. **07**

Sr. no.	Outlook	Temperature	Humidity	Wind	Play
1	Sunny	hot	high	FALSE	No
2	Sunny	hot	high	TRUE	No
3	Overcast	hot	high	FALSE	Yes
4	Rain	mild	high	FALSE	Yes
5	Rain	cool	normal	FALSE	Yes
6	Rain	cool	normal	TRUE	No
7	Overcast	cool	normal	TRUE	Yes
8	Sunny	mild	high	FALSE	No
9	Sunny	cool	normal	FALSE	Yes
10	Rain	mild	normal	FALSE	Yes
11	Sunny	mild	normal	TRUE	Yes
12	Overcast	mild	high	TRUE	Yes
13	Overcast	hot	normal	FALSE	Yes
14	Rain	mild	high	TRUE	No

- Q.4** (a) Do feature wise comparison between OLAP and OLTP. **03**
- (b) Define data cube and explain 3 operations on it. **04**
- (c) Define linear and nonlinear regression using figures. Calculate the value of Y for X=100 based on Linear regression prediction method. **07**

X	Y
4	390
9	580
10	650
14	730
4	410
7	530
12	600
22	790
1	350
3	400
8	590
11	640

5	450
6	520
10	690
11	690
16	770
13	700
13	730
10	640

OR

- Q.4** (a) Explain Spatial mining using example. **03**
 (b) Calculate the weights using neural network single layer perceptron model. Three inputs are x_0, x_1, x_2 , bias and weights are as follows:
 $w_1(0) = 30, w_2(0) = 300$
 $b(0) = 50, \eta = 0.01, x_0 = +1$
 Activation function is :
 $\text{sgn}(x) = +1, \text{ if } x \geq 0$
 $\text{sgn}(x) = -1, \text{ if } x < 0$
 (a) Calculate x_2 for $x_1 = 100$ and $x_2 = 200$.
 (b) For bias $b(0) = -1230$ recalculate the weights w_1 and w_2 .
 (c) How data Mining is useful for Business Intelligence applications viz. Balanced Scorecard, Fraud Detection, Clickstream Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance and CRM **07**
- Q.5** (a) Explain text mining using example. **03**
 (b) Explain big data and big data analytics. Explain key roles and their responsibilities for successful analytic project. **04**
 (c) Calculate 2 clusters using k-means cluster algorithm. For finding the distance use euclidian distance. **07**

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

Assume mean1 as subject1 and mean2 as subject4

OR

- Q.5** (a) Explain web mining using example. **03**
 (b) Explain Hadoop architecture using figure. **04**
 (c) Explain mapreduce. Explain any example using mapreduce. **07**