No. of Printed Pages : 8

MINE-022

M.Tech. IN ADVANCED INFORMATION TECHNOLOGY – SOFTWARE TECHNOLOGY (MTECHST)

00013 **Term-End Examination**

June, 2015

MINE-022 : BUSINESS INTELLIGENCE

Time : 3 hours *Maximum Marks : 100*

Instructions to the Candidates :

(i) *Section I is compulsory and carries 30 marks.*

(ii) *Section II carries 70 marks. Answer* ***any five*** *questions.*

(iii) *Assume suitable data wherever required.*

(iv) *Draw suitable sketches wherever required.*

(v) *Italicized figures to the right indicate maximum marks.*

SECTION I

1. As a part of a promotional campaign, the marketing manager of "Dilshan Retails" wishes to identify the potential buyers to send promotional catalog offers. The manager uses predictive modeling tools to find the potential customers for promotional offers. He takes help from one of the statisticians working in the same company. The objective of the selection is to

"Maximize profit with cost". The variables for the selection are as follows :

Variables	Variables
Purchase(c,2)	Total Returns
Dollars Spent	Mens Apparel
Yearly Income	Home Furniture
Home Value	Lamps Purchase
Order Frequency	Linens Purchase
Recency	Blankets Purchase
Married(c,2)	Towels Purchase
Name Prefix(c,4)	Outdoor Product
Age	Coats Purchase
Sex(c,2)	Ladies Coats
Telemarket Ind.(c,2)	Ladies Apparel
Rents Apartment	His/Her Apparel
Occupied <1 Year	Jewelry Purchase
Domestic Product	Date 1st Order
Apparel Purchase	Telemarket Order(c,5)
Leisure Product	Account Number(c,128)
Luxury Items(c,2)	State Code(c,55)
Kitchen Product	Race(c,5)
Dishes Purchase	Heating Type(c,4)
Flatware Purchase	Number of Cars(c,4)
Total Dining (kitch+dish+flat)	Number of Kids
Promo:1-7 Months	Travel Time
Promo: 8-13 Months	Education Level(c,4)
$Value per Mailing	Job Category
Country Code	
	Note: 'c' indicates class levels)

Answer the following questions :

(a) What variables (input and target) are important for potential customer selection ? Which method do you prefer for variable selection ? *3+3=6*

(b) Identify the suitable predictive analytics technique (model) and write reasons for choosing the technique. *6*

(c) How do you apply the model using SAS/SEMMA methodology ? *8*

2. A sample of 20 observations was taken to conduct the Regression analysis to find the value of weight when height is given. Sample data having variables names, weight, height and age is as follows :

Alfred, 69·0, 112·5, 14

Alice, 56·5, 84·0, 13

Barbara, 65·3, 98·0, 13

Carol, 62·8, 102·5, 14

Henry, 63·5, 102·5, 14

After applying simple linear regression, the following results were produced in SAS system.

Simple Linear Regression

The REG Procedure

Model : MODEL 1

Dependent Variable : Weight

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	7193.24912	7193.24912	57.08	< 0.0001
Error	17	2142.48772	126.02869		
Corrected Total	18	9335.73684			

Root MSE	11.22625	R Square	0.7705
Dependent Mean	100.02632	Adj R-Sq	0.7570
Coeff. Var	11.22330		

Answer the following short answer questions : *5+5=10*

(a) Write SAS code to generate the above result.

(b) Write conclusions for the results shown.

SECTION II

3. (a) What is Data Pre-processing ? Is pre-processing essential for Business intelligence applications ? Write any three consequences if data pre-processing is not done. *2+2+3=7*

(b) What are the measures of Central Tendency ? How do measures of Central Tendency help the business analyst in Business Intelligence Applications ? *3+4=7*

4. (a) What is Business Analytics (BA) ? How is BA useful to analysts ? *3+3=6*

(b) Apply any one of the decision tree algorithms for the following bank loan applicants data. Explain step-by-step. *8*

Home	Marital status	Gender	Employed	Credit Rating	Risk Class
Yes	Yes	Male	Yes	A	B
No	No	Female	Yes	A	A
Yes	Yes	Female	Yes	B	C
Yes	No	Male	No	B	B
No	Yes	Female	Yes	B	C
No	No	Female	Yes	B	A
No	No	Male	No	B	B
Yes	No	Female	Yes	A	A
No	Yes	Female	Yes	A	C
Yes	Yes	Female	Yes	A	C

5. (a) Find the association rules with minimum support and minimum confidence 50%, using Apriori algorithm for the following table of telephone faceplate purchases. The numeric 1 (one) indicates the presence of an item in the transaction and numeric 0 (zero) indicates the absence of transaction. *9*

Transaction_id	Red	White	Blue	Orange	Green	Yellow
1	1	1	0	0	1	0
2	0	1	0	1	0	0
3	0	1	1	0	0	0
4	1	1	0	1	0	0
5	1	0	1	0	0	0
6	0	1	1	0	0	0
7	1	0	1	0	0	0
8	1	1	1	0	1	0
9	1	1	1	0	0	0
10	0	0	0	0	0	1

(b) Explain the suitable performance measures for the above stated problem (Q.5(a)). *5*

6. (a) Apply K-means Clustering algorithm for the following student data to get suitable clusters. 9

Student	Age	Subject1	Subject2	Subject3
S1	18	73	75	57
S2	18	79	85	75
S3	23	70	70	52
S4	20	55	55	55
S5	22	85	86	87
S6	19	91	90	85
S7	20	70	65	60
S8	21	53	56	59
S9	19	82	82	60
S10	47	75	76	77

(b) Explain the most widely used measures for measuring distance between clusters. 5

7. (a) CRM (Customer Relationship Management) applications require large amount of customer data (past data, current and external data) to give better Business Intelligence. Why ? 5

(b) Suppose the data for analysis includes the attribute age. The age values for data tuples are :

13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

(i) Find mean, median and mode. 3

(ii) What is midrange of the data ? *3*

(iii) Find first quartile (Q1) and third quartile (Q3) of the data. *3*

8. The following table shows the midterm and final exam grades for students in a database course. *6+6+2=14*

X-midterm	Y-Final exam
72	84
50	63
81	77
74	78
94	90
86	75
59	49
83	79
65	77
33	52
88	74
81	90

(a) Plot the data. Do X and Y seem to have a linear relationship ?

(b) Use the method of least squares to find an equation for the prediction of a student's final exam grade based on the student's midterm grade in the course.

(c) Predict the grade in final exam of a student who received 86 on the midterm exam.