

The centroids of the three clusters in terms of ('alcohol', 'alkalinity of ash') are (10.88, 12.83), (12.37, 20.46) and (14.44, 28.71) respectively. To which cluster, a new wine (say, wine 21) with ('alcohol', 'alkalinity of ash') as (12,12) shall be assigned based on Euclidean Distance measure. Illustrate with computations. **5 Marks**

5. [i] The following table gives the demand (Kg) and price (Rs/Kg) figures for a commodity for 6 days. Fit the regression  $P = a + b \cdot D$  (Where P: Price, D: Demand; a and b are constants). Present the calculations involved.

Days	1	2	3	4	5	6
Price	22	30	25	20	15	8
Demand	10	12	15	20	23	28

Estimate the price when demand is 22? **5 Marks**

[ii] You have been asked to assess performance of a B-School. Keeping in view various aspects of a B-school performance, what are the different types of data which you consider important to collect for this. **5 Marks**

6. Write short notes on any two of the following: **5X2 =10 Marks**

- Difference among descriptive, predictive and prescriptive analytics using suitable illustrations
- Type I and Type II errors in Hypothesis Testing
- Dimensions of data quality with illustrations
- Key stages in a business analytics project

-END-

Total No. of Pages 4

THIRD SEMESTER

SUPPLEMENTARY EXAMINATION

MGT-31 BUSINESS ANALYTICS

Time: 3:00 Hours

Roll No. ....

MBA

FEB-2019

Max. Marks: 50

Note: Attempt five questions in all. Section A is compulsory. Assume suitable missing data, if any. Use of calculator is allowed.

SECTION A

1. You are required to write True/False, choose correct option(s) or solve or fill in the blanks. Each of parts i to x is of 1 mark.

10 Marks

- The range for R square is :  
(a) 0 to 1 (b) 0 to  $\infty$  (c)  $-\infty$  to  $\infty$  (d) -1 to 1
- If there is multi-collinearity in a regression model, the model can be used for making good predictions (True/False)
- Durbin-Watson statistic is used for testing presence of -----  
-----
- The difference in means of two normally distributed populations can be tested using -----  
(t-test / Chi-square test / r-test)
- The clusters should be so formed that the objects within a cluster are homogeneous (True/False)
- Assignment problem is an example of -----analytics.  
(Descriptive or Prescriptive or Predictive)

P.T.O.

P.T.O.

vii. Every feasible solution to a linear programming problem  $P$  is also an optimal solution (True/False)

viii. In a logistic regression, the logit function is  $\text{Log} [ P(Y=1) / ( 1 - P(Y=1) ) ] = -4 + 0.25X$ . Then the equation for logit function  $\text{Log} [ P(Y=0) / ( 1 - P(Y=0) ) ]$  is (write the steps of solution):

a.  $-0.25 + 4X$  (b)  $0.25 - 4X$  (c)  $4 - 0.25X$  (d)  $4 + 0.25X$

ix. In ANOVA test, the hypothesis of no difference in population means is tested using:

(a) Omnibus test (b) Wald's test (c) F-test (d) z-test

x. Cluster algorithms are ----- (supervised or unsupervised) learning algorithms.

### SECTION B

2. [i] The liking for tourist places in respect of two persons is as follows:

Person 1 = {Mukteshwar, Mumbai, Dharamsala, Musoorie, Kanayakumari, Srinagar, Shillong}

Person 2 = {Mukteshwar, Musoorie, Goa, Dharamsala, Shillong, Srinagar}

The places which are not included in a set are disliked by the respective persons. Considering liking as 1 and disliking as 0,

P.T.O.

present the above data in matrix form. Compute Jaccard coefficient and comment on similarity of the two persons. **5 Marks**

[ii] Movie ratings given by two customers on a 5 point scale are as follows:

Customer	Ghajini	Dilwale	Uri	Guide	Don	Hero
Customer 1	5	1	5	5	1	1
Customer 2	1	5	1	1	5	5

Compute Cosine coefficient and comment on the similarity of the customer ratings. **5 Marks**

3. [i] What are the challenges which are likely to be faced during data pre-processing stage of an analytics project. Explain with suitable illustrations. **5 Marks**

[ii] The following are sample tuples from employee database of a startup:

Name	Gender	Date of Birth	Joining Date	Skill	Experience (years)	Performance Score	Rank
Amit	M	01-07-1990	11-07-2018	IT	3	90	5
Smita	F	01-09-1991	03-09-2018	Design	1	85	3
---	---	---	---	---	---	---	---

Specify attribute type for each attribute. What type of transformation is possible on each attribute type? **5 Marks**

4. [i] What is the difference between training and testing data sets? **5 Marks**

[ii] Twenty (20) types of wines (labeled as 1 to 20) are clustered based on their 'alcohol' and 'alkalinity of ash' contents as follows:

Cluster 1 - {2,6,7,10,11,15,17,18}, Cluster 2 - {3,4,8,13,20}

Cluster 3- {1,5,9,12,14,16,19}

P.T.O.