

Registration No:

--	--	--	--	--	--	--	--	--	--

Total Number of Pages : 1

M.TECH

M.TECH 2ND SEMESTER (AR 18) REGULAR EXAMINATIONS, APRIL/MAY 2019

DATA PREPARATION AND ANALYSIS

Branch: CSE, Subject Code:MCSPE2031

Time: 3 Hours

Max Marks : 70

PART-A

(10 X 2=20 MARKS)

1. Answer the following questions.

- List the data preparation steps.
- What are the reasons for which an organization collects data?
- What is data discretization?
- Distinguish between a Population and a Sample.
- Define outlier?
- How do you measure the variation in data?
- What is clustering?
- Differentiate supervised and unsupervised learning.
- What is predictive analysis?
- Define Entropy?

PART-B

(5 X 10=50 MARKS)

Answer any five questions from the following.

- Q2.a) What are the different categories, based on scales of measurement, we classify the data? [5]
 b) What is Data Transformation? List the common data Normalization steps performed on data. [5]
- Q3.a) List and describe the steps carried out during data preparation. [5]
 b) Use Min-Max normalization by setting min=0 and max=1, to normalize the following data. [5]
 200, 300, 400, 600, 1000.
- Q4.a) What is a contingency table? What insight does it provide about the dataset. [5]
 b) List the different ways of visualizing information using graphs. [5]
- Q5.a) Distinguish between Inferential Statistics and Comparative Statistics. [5]
 b) Justify how outliers affect the performance of k-Means clustering algorithm? [5]
- Q6. a) What is a Decision Tree? Taking an example data set, show how the splitting criteria is defined. [5]
 b) Describe briefly about the statistical measures for quantifying the symmetry or skewness in the data. [5]
- Q7. Suppose that a hospital tested the age and body fat data for 18 randomly selected adults with the following results:

<i>age</i>	23	23	27	27	39	41	47	49	50
<i>%fat</i>	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
<i>age</i>	52	54	54	56	57	58	58	60	61
<i>%fat</i>	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

- Calculate the mean, median, and standard deviation of age and %fat [5]
 - Find out the correlation among these two attributes. [5]
- Q8. Write short notes on:
- Chi-square test [5]
 - Associative rules [5]

==0==