

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



GIET UNIVERSITY, GUNUPUR – 765022
M. Tech (Second Semester Examinations) – October' 2021
MPECS2031 – DATA PREPARATION AND ANALYSIS
(C.S.E)

Time: 2 hrs

Maximum: 50 Marks

(The figures in the right hand margin indicate marks)

PART – AQ.1. Answer *ALL* questions

(2 x 10 = 20)

- What are the reasons for which an organization collects data?
- What is Scalability?
- Identify the real time issues in data preparation?
- Distinguish between a Population and a Sample.
- What is segmentation?
- Why Data cleaning is an essential pre-processing step in data analysis?
- What is over fitting?
- Differentiate between supervised and unsupervised learning.
- What do you mean by data visualization?
- What is a Box plot?

PART – B**(6 x 5 = 30 Marks)**Answer *ANY FIVE* questions

Marks

- What are the roles and responsibilities of a Business analyst and Data mining consultant in the implementation of data analysis project. (6)
- Discuss in detail about the process of converting text to numbers. (6)
- Find in dealing with missing values in heterogeneous data using k-nearest neighbours? (6)
- Differentiate between (6)
 - Descriptive and Inferential Statistics.
 - heterogeneous and homogeneous data sets
- What is a Summary Table? What insight does it provide about the dataset? (6)
- Describe the merits and shortcomings of K-Nearest Neighbour method of prediction (6)
- Write short notes on (6)
 - Gradient Descent
 - Linear Discriminant Analysis

--- End of Paper ---