School of University Polytechnic
Diploma in Computer Science and Engineering
Mid Term Examination - Nov 2023

Duration : 90 Minutes
Max Marks : 50

Sem III - N1DE301B - Introduction to Data Science

General Instructions
Answer to the specific question asked
Draw neat, labelled diagrams wherever necessary
Approved data hand books are allowed subject to verification by the Invigilator

1) Explain data cleaning. — K2 (2)

2) Define data reduction and its significance in Data Science — K1 (3)

3) Discuss the potential challenges or limitations of data discretization when dealing with highly skewed distributions. — K2 (4)

4) Discuss the advantages and disadvantages of using clustering algorithms for instance reduction. — K2 (6)

5) Determine the examples of both structured and unstructured data and explain how data cleaning processes may differ for each type. — K3 (6)

6) Illustrate various data collection tools. — K3 (9)

7) Critically analyze the advantages and limitations of different data collection strategies such as surveys, sensor data, and social media mining. — K4 (8)

8) Analyze the main factors that have contributed to the evolution of Data Science over the years. — K4 (12)

OR

Analyze the potential security risks associated with collecting and analyzing sensitive data, and propose strategies to mitigate these risks. — K4 (12)