

ADMISSION NUMBER											

**School of Business**  
**Bachelor of Business Administration**  
**Mid Term Examination - May 2024**

**Duration : 90 Minutes**  
**Max Marks : 50**

**Sem IV - D1UF401T - Data Mining**

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) Illustrate the challenges associated with data mining, such as data quality issues, scalability, and interpretability of results. How can these challenges be addressed? K3 (6)
- 2) Find potential outliers in a dataset using the interquartile range (IQR) rule. Interpret your findings. Data: Data set: {5, 7, 10, 15, 19, 20, 21, 22, 23, 23, 23, 23, 24, 24} K3 (9)
- 3) Explain the two-sigma range method for detecting outliers. K4 (4)
- 4) Analyse the role of feature selection in predictive analytics. K4 (8)
- 5) How, within the Knowledge Discovery in Databases (KDD) framework, does the systematic analysis of historical workplace incidents enable knowledge discovery? Analyse how this analytical approach informs decision-making by assessing the likelihood of future risks in specific operational areas for the company. regarding the likelihood of future risks in specific operational areas for the company? K4 (8)
- 6) What are the common approaches for dealing with outliers in data cleaning? Compare them and describe which is more useful under which situation. K5 (5)
- 7) A retail company wants to identify customer segments based on their purchase history. The company has a large dataset containing transaction data, including customer demographics, product categories purchased, and purchase amounts. Recommend which data mining technique could be used to identify customer segments with distinct purchasing behaviors. Describe its benefits over other alternative techniques. K5 (10)