

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

## School of Computing Science and Engineering

Bachelor of Science in Computer Science

Semester End Examination - May 2024

Duration : 180 Minutes

Max Marks : 100

### Sem VI - E1UP603B - Data Analysis and Handling

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) Define the aggregative statistics methods used with DataFrame in Pandas? K1 (2)
- 2) How do you assess the robustness of a predictive model, and why is it important? Explain. K2 (4)
- 3) Explain the concept of regularization in predictive modeling, and also why is it used. K2 (6)
- 4) Examine a financial dataset to identify fraudulent activities or anomalies in transaction patterns, and apply data-driven approaches to enhance fraud detection and prevention. K3 (9)
- 5) Analyze a real-world dataset from the healthcare domain and apply a data-driven solutions to improve patient outcomes. K3 (9)
- 6) Recommend a machine learning model for sentiment analysis of customer reviews in the hospitality industry. K5 (10)
- 7) In the realm of environmental science, analyze climate data to analyze the impact of human activities on global warming trends and propose mitigation strategies. K4 (12)
- 8) Determine the use of data storytelling in journalism and media to communicate complex socio-political issues and engage diverse audiences. K5 (15)
- 9) Determine the scalability and performance of interactive visualization techniques for handling large and streaming datasets in real-time analytics applications. K5 (15)
- 10) Develop an advanced algorithm for automatic anomaly detection in streaming data using real-time visualization feedback to adaptively adjust detection thresholds and improve anomaly detection performance. K6 (18)