M.C.S.E. FIRST YEAR SECOND SEMESTER EXAMINATION-2025
(Internet of Things)
DATA SCIENCE AND VISUALIZATION

Time: 3 hours **Full Marks: 100**

---

1. A (i) Explain the invertibility condition of MA(1) and its importance in model estimation. Determine whether the following MA(1) process is invertible: $X_t = \mathbf{a}_t + \beta_1 \mathbf{a}_{t-1}$ for $\beta_1 = -1.5$.
(ii) Ten successive observations on a stationary time series: 0.6, 1.8, 2.2, 0.5, 1.6, 0.7, 0.9, 1.3, 2.5, 0.3. Compute autocorrelation coefficients r_1.
(iii)For the model $(1-0.3B)(1-B)X_t = (1-2.7B)\mathbf{a}_t$, find p, d, q and express it as ARIMA (p, d, q). Determine whether the process is stationary and invertible.

6 + 8 + 6

OR

1.B.(i) Suppose a time series has increasing variance and a nonlinear trend. Explain how to preprocess the data using transformations (log, square root, etc.) before applying ARIMA modeling. Discuss how these transformations affect the interpretation of AR, MA, and ARIMA forecasts.
(ii) Given a time series: 12, 15, 14, 16, 18, 17, assume it follows an AR(1) model: $X_t = \boldsymbol{\alpha} X_{t-1} + \mathbf{a}_t$, where $\mathbf{a}_t$ is white noise. Estimate the coefficient $\boldsymbol{\alpha}$ and predict the next value X_7.

10 + 10

2.A.(i) Describe the process of creating a boxplot. What insights can be drawn from it about the distribution of data?
(ii) What are the potential pitfalls of using pie charts or multiple donut charts for comparing multiple categorical variables? Suggest better alternatives.
(iii) What are the principles of good visual design in data visualization, and how do they help balance aesthetics with clarity and accuracy? Explain using real or hypothetical visual examples.
(iv) Explain how a scatter plot and correlation matrix are used to explore relationships between variables. Illustrate with an example.
(v) Describe how treemaps can be used to visualize proportions in hierarchical data. What are the benefits and drawbacks of using treemaps over traditional charts?

4 × 5

OR

2.B.(i)Discuss the importance of designing color-blind-friendly data visualizations. What are the consequences of failing to accommodate color-blind users?
(ii) Discuss the role of axes, gridlines, and projection in ensuring clarity in both 2D and 3D plots. How can poor design affect data comprehension?
(iii) Explain how excessive decorative elements (such as 3D effects, shadows, gradients, and images) can interfere with the effectiveness of data visualizations. When is decoration helpful versus harmful?
(iv) Define whitespace in the context of data visualization. How does strategic use of whitespace improve readability, hierarchy, and visual appeal?
(v) Explain the use of ridgeline plots in visualizing multiple distributions. What are the visual and interpretative advantages of this technique?

4 × 5

3.A.(i) Describe an end-to-end data science pipeline that includes data collection, cleaning, preprocessing, handling class imbalance, modeling, and evaluation. Use a real-world use case like fraud detection or disease diagnosis to illustrate your workflow.
(ii) Discuss the role of digital tools and technologies in modern data collection. Examine the impact of mobile apps, web-based forms, APIs, and IoT sensors in automating and enhancing data collection. Include examples from different industries. 10 + 10

[Turn over

OR

3.B(i) Class imbalance is a common problem in datasets such as credit card fraud detection or cancer prediction. How do class imbalance affect model accuracy, precision, and recall?

(ii) What are the benefits in terms of reproducibility, maintainability, and error reduction?

(iii) Explain how outliers can affect model performance. Discuss a method to detect and treat outliers.

8 + 6 + 6

4.A.(i) A training institute wants to prove that their new course improves performance. Student test scores before and after the course are:

Before: 60, 62, 61, 59, 58

After: 65, 67, 64, 66, 68

a) Perform a paired t-test.

b) Interpret the effectiveness of the training program.

(ii) The weights (in kg) of a random sample of 10 boxes are:

10, 12, 11, 13, 12, 14, 11, 12, 13, 12

a) Compute the 95% confidence interval for the population mean.

b) Interpret the result and explain what it implies for decision-making in quality control.

10 + 10

OR

4.B.(i) Describe the use of the t-test in statistical inference. Differentiate between one-sample, independent-sample, and paired-sample t-tests, with examples for each.

(ii) How do you determine if a dataset is normally distributed? Explain the importance of normality in statistical inference

(iii) Discuss the impact of sample size and sampling methods on statistical inference.

7 + 6 + 7

5.A.(i)Explain the role of data fusion in Internet of Things (IoT) sensor networks. How does fusion improve data quality, reduce redundancy, and enhance energy efficiency? Provide examples from smart cities or agriculture.

(ii) A robot estimates its position using two range sensors:

Prior: P(Object at 5m)=0.4, P(Object at 10m)=0.6

Sensor 1 reports 5m with P(S1=5m|Obj=5m)=0.9, P(S1=5m|Obj=10m)=0.3

Sensor 2 reports 5m with P(S2=5m|Obj=5m)=0.85, P(S2=5m|Obj=10m)=0.25

Use Bayesian inference to calculate the posterior probability that the object is at 5 meters.

(iii) Two sensors provide conflicting information on object type: Θ={Red, Blue}

Sensor 1: m1(Red)=0.9, m1(Blue)=0.05, m1(Red∪Blue)=0.05

Sensor 2: m2(Red)=0.1, m2(Blue)=0.85, m2(Red∪Blue)=0.05

Use Dempster's Rule to compute the combined belief.

8 + 6 + 6

OR

5.B.(i) Define spatial interpolation. Why is interpolation necessary in spatial data analysis, and what are its typical applications?

(ii) Describe the basic principle of Inverse Distance Weighting (IDW) interpolation. How is it different from simpler interpolation methods like Kriging?

(iii) Suppose you have a predicted estimate with variance 6, and a sensor measurement with variance 3. Calculate the Kalman gain. What does this value tell you about the trust given to the measurement?

(iv) A moving object's state is position and velocity:

$$x=\begin{bmatrix} \text{position} \\ \text{velocity} \end{bmatrix}=\begin{bmatrix} 0 \\ 1 \end{bmatrix},\ P=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The transition model is:

$$A=\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

Assume measurement matrix H = [1 0], and measurement of position is 1.2 with noise variance 0.2. Perform prediction and update step of Kalman Filtering. Provide updated state and covariance matrix.