



23MCT305 – DATA ANALYTICS IN AUTOMATION SYSTEM
TECHNICAL PUZZLES

UNIT III: DATA ANALYTICS (CO3)

Puzzle No.	Puzzle Description	Expected Approach / Key Concept	Bloom's Level	CO
1	You have a dataset to predict machine failure (Yes/No). Explain why you would prefer logistic regression over linear regression, and then why you might switch to a decision tree instead. Provide one scenario each.	Logistic for probability; Decision tree for interpretability/non-linearity	Analyze	CO3
2	A decision tree splits first on temperature $> 80^{\circ}\text{C}$ (Gini drops from 0.5 to 0.3), then on vibration > 5 mm/s. Manually compute the information gain if entropy was initially 1.0 and becomes 0.4 after split (assume balanced split). Show steps.	Information Gain = $1.0 - 0.4 = 0.6$	Apply	CO3
3	Your linear regression model gives perfect fit on training data but poor on test data. Diagnose the problem and suggest two remedies using only techniques from Unit III.	Overfitting; remedies: pruning (if tree), regularization (for regression)	Evaluate	CO3
4	Rank linear regression, logistic regression, and decision tree in terms of interpretability for a shop-floor engineer who needs to explain model predictions to operators. Justify the ranking.	Decision tree $>$ Logistic $>$ Linear	Evaluate	CO3
5	Design a hybrid approach: use linear regression for trend prediction and decision tree for anomaly classification on the same sensor dataset. Describe how the outputs of one can feed into the other.	Residuals from linear \rightarrow features for decision tree	Create	CO3