

## II SEMESTER M.TECH. (INDUSTRIAL AUTOMATION AND ROBOTICS) END SEMESTER EXAMINATION- 7<sup>st</sup> MAY 2024

SUBJECT: Data Analytics for Automation Subject		ct Co	04				
Time: 3 Hour	Exam time: 9:30-12:30 am	MAX. I	MAR	KS: 5(	)		
	Instructions to Candidates:						
<ul><li>✤ Answer AI</li></ul>	L the questions.						
<ul><li>✤ Assume su</li></ul>	itablt any data that is not provided.						
			М	CO	РО	LO	BI
Calculate the posterior p given a positive test resu	probability of a patient having a certain alt, based on the following information	ı disease :	2	1	1,3	1,2	3
<ul><li>Prior Probability</li><li>Likelihood: 60%</li><li>Positive Test Res</li></ul>	: 25% sult: 40%						
In a study examining the and sales revenue for a from various regions ove	e relationship between advertising exp particular product, researchers collec er a period of time.	enditure ted data	3	3	5	4	3
After conducting regrese exhibit unequal varian expenditure. Identify the regression analysis.	ssion analysis, they found that the name across different levels of advice problem observed in the residuals	residuals vertising s of the					
In the context of regress a widely used metric to data. Consider the follow	ion analysis, the coefficient of determine assess the goodness of fit of a mode wing case study:	nation is el to the	5	2	1,3	1,2	5
A research team is conbetween a set of independent and income) and a depsample of individuals. participants and fitted a health score based on the	nducting a study to examine the relatendent variables (such as age, education bendent variable (e.g., health score) a . The team has collected data from multiple linear regression model to pro- e independent variables.	ationship on level, among a om 100 edict the					
Discuss the need for an context of this case. Cor sample size, and model of	n adjusted coefficient of determinationsider factors such as the number of proceeding of the second se	n in the edictors,					

4.	Assuming a binomial distribution, calculate the probability of getting exactly 3 heads in 5 coin tosses, assuming a fair coin.	g <b>2</b>	2	1,3	1,2	3
5.	In a sociological study investigating social media addiction among university students, researchers utilize a method where participants are recruited based on referrals from their peers.	g 4	2	1,3	1,2	4
	Describe the recruitment process employed in this study and discuss the potential implications of this approach. Identify both advantages and disadvantages of using this recruitment method in the context of the study.	5 5 f				
6.	In the context of distributed systems design, the trade-offs between consistency, availability, and partition tolerance, as described by Brewer's CAP Theorem, play a crucial role.	<b>4</b>	1	1,3	1,2	4
	<ul> <li>a) Analyze the core principles of consistency, availability, and partition tolerance in distributed systems, as outlined by Brewer's CAP Theorem.</li> </ul>	1				
	b) Justify the importance of understanding and considering Brewer's CAP Theorem in the design and implementation of distributed systems, highlighting its impact on system behavior and performance.					
7.	Calculate the adjusted R-squared value for a multiple regression mode with 100 data points and 5 predictor variables	2	2	1,3	1,2	4
8.	Suppose the life expectancy of people in a population is normally distributed with a standard deviation of 1.	3	2	1,3	1,2	3
	You go out and sample 45 people from this population and obtain a mean life expectancy of 88.51 and a standard deviation of 1.0815.	l				
	Using an alpha value of $= 0.05$ , is this observed mean signicantly different than an expected life expectancy of 89? (critical z-value is approximately ±1.96)					
9.	Explain the role of each element within the Enterprise Performance Life Cycle Framework in optimizing organizational performance	e 5	3	5	4	4
10.	Consider the output of a test using a Multiple Linear Regression.	2	3	5	4	4
	Residuals: 1 2 3 4 5 -2.65089 0.06756 -1.37741 0.49274 3.46800					
	Coefficients: Estimate Std. Error t value Pr(> t ) (Intercept) 23.55802 4.75947 4.961 0.0383 * distance -0.04647 0.03247 -1.431 0.2887 emission -0.15838 0.46144 -0.333 0.7643					
	emilission -0.15820 0.40144 -0.345 0.7045  Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
	Residual standard error: 3.256 on 2 degrees of freedom Multiple R-squared: 0.5691, Adjusted R-squared: 0.1383 F-statistic: 1.321 on 2 and 2 DF, p-value: 0.4309					

Comment based on p-value which of the variable has the least significance and the rationale behind it.

11. 4 1.3 1.2. 3 3 Outline the sequential stages of the Data Mining Process, emphasizing the primary tasks involved at each stage. 3 1.3 1.2, 4 12. Suppose you are working as a data scientist for a financial institution 5 4 that wants to develop a model to predict customer churn. The 3 institution provides various banking services, including savings accounts, credit cards, loans, and investment products. The dataset contains historical information about customer demographics, transaction history, account balances, customer interactions, and whether or not each customer churned within a specified period. As a part of your analysis, you decide to explore ensemble learning techniques to improve the predictive performance of your models. Considering the characteristics of the dataset and the business context, which ensemble learning techniquewould you recommend using? Provide a detailed explanation to justify your choice. 13. 2 3 5 3 4 Enlist four characteritistics of a good sample. 3 4 1.3 1.2. 5 14. Create a SQL query that effectively utilizes the JOIN statement to produce a result set showing each employee's name alongside their 3 corresponding department name. Ensure the query is optimized for performance and readability.

Table 14 a

EmployeeID	Name	DepartmentID		
1	John	101		
2	Alice	102		
3	Bob	103		

Table 14 b

DepartmentID	DepartmentName
101	Sales
102	Marketing
103	Finance

15

Given a case study of a rapidly growing e-commerce platform experiencing performance bottlenecks due to increased user activity, evaluate the suitability of sharding as a solution for scaling their database infrastructure.

Consider factors such as data distribution, query performance, and operational complexity in your analysis

3.5 1.2, 5

3,17

5

5