



Enrollment No.....

Faculty of Engineering
End Sem (Odd) Examination Dec-2022
IT3ED03 Data Analytics
 Programme: B.Tech. Branch/Specialisation: IT

Duration: 3 Hrs.**Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d.

- Q.1 i. In which of the following learning, output label is known- **1**
 (a) Unsupervised (b) Supervised
 (c) Both (a) & (b) (d) None of these
- ii. The goal of using data mining technique is to- **1**
 (a) Detect outliers (b) Detect noise
 (c) Detect patterns (d) None of these
- iii. In data pre-processing, outliers represent- **1**
 (a) Abnormal values (b) Mean value
 (c) Mode value (d) None of these
- iv. ETL stands for- **1**
 (a) Easy-to-load (b) Easy-transform-load
 (c) Extract, transform, load (d) None of these
- v. In dimension reduction, dimension represents **1**
 (a) Value (b) Variable (c) Key (d) None of these
- vi. PCA technique is basically used to- **1**
 (a) Emphasize variation (b) Easy visualization
 (c) Data cleaning (d) None of these
- vii. The level-C in confidence interval gives- **1**
 (a) Variation (b) Mean (c) Probability (d) Deviation
- viii. The probability of committing Type-I error is called- **1**
 (a) Beta level (b) Significance level
 (c) Power level (d) None of these
- ix. Variance is a measure of how the data values is dispersed around the **1**
 (a) Average (b) Mean (c) Deviation (d) Mode

P.T.O.

- x. Z-test is performed when the- **1**
 (a) Variations are known (b) Sample size is less
 (c) Mean is known (d) None of these
- Q.2 i. Explain the importance of data mining. **2**
 ii. Differentiate between supervised learning and unsupervised learning. **3**
 iii. Compare parametric model with non-parametric model. **5**
 OR iv. Categorize various data analytics methodologies. **5**
- Q.3 i. What do you mean by feature creation? **2**
 ii. Discuss Anscombe's quartet with an example. **3**
 iii. What are the various strategies used for data transformation? **5**
 OR iv. Discuss the various techniques of data audit to deal with missing values and outliers processing. **5**
- Q.4 i. Explain the issues associated with data preparation. **2**
 ii. What is the significance of dimension reduction? **3**
 iii. Discuss the Principle Component Analysis (PCA) technique with taking the example of 2D datasets. **5**
 OR iv. Apply the factor analysis techniques on questionnaire data sets. **5**
- Q.5 i. What is the difference between estimation and prediction? **2**
 ii. Explain any model of performing statistical inference. **3**
 iii. Discuss the various steps involved for calculating the confidence interval with an example. **5**
 OR iv. What do you mean by the term "Null Hypothesis"? Also discuss Type-I and Type-II error associated with it. **5**
- Q.6 i. Explain the importance of multivariate statistics. **2**
 ii. Compare two-sample t-test with two-sample z-test. **3**
 iii. Discuss the various steps for performing Chi-square test for homogeneity. **5**
 OR iv. Write down the steps for calculating the two-sample z-test for two population proportions. **5**

Marking Scheme
IT3ED03 Data Analytics

	(a) Unsupervised (b) Supervised (c) Both (a) & (b) (d) None of these	
ii	The goal of using Data Mining Technique is to	1
	(c) Detect Patterns	
iii	In data pre-processing, outliers represents	1
	(a) abnormal values	
iv	ETL stands for	1
	(c) Extract,transform,load	
v	In Dimension Reduction, dimension represents	1
	(b) variable	
vi	PCA technique is basically used to	1
	(a) emphasize variation	
vii	The level-C in confidence interval gives	1
	(c) Probability	
viii	The probability of committing Type-I error is called	1
	(b) Significance Level	
ix	Variance is a measure of how the data values is dispersed around the	1
	(b) Mean	
x	Z-test is performed when the	1
	(a) Variations are known	

Q.2	i.	Explain the importance of Data Mining?	3
		For each importance -1 mark x 3=3 marks	
	ii	Differentiate between supervised learning and unsupervised learning?	2
		For each difference -1 mark x 2=2 marks	
	iii	Compare Parametric model with Non-Parametric model?	5
		For each comparison -1 mark x 5=5 marks	
OR	iv	Categorize various Data Analytics methodologies?	5
		For each category -1 mark x 5=5 marks	

Q.3	i.	What do you mean by feature creation?	2
		Definition -2 marks	
	ii	Discuss Anscombe's Quartet with an example?	3
		Definition -1 mark	

		Example	-2 marks
	iii	What are the various strategies used for Data transformation?	5
		For each strategy -1 mark x 5=5 marks	
OR	iv	Discuss the various techniques of data audit to deal with missing values and outliers processing?	5
		For each missing value technique -1 mark x 2=2 marks	
		For each outliers processing techniques -1 mark x 3=3 marks	
Q.4	i.	Explain the issues associated with data preparation?	2
		For each issue -1 mark x 2=2 marks	
	ii	What is the significance of Dimension Reduction?	3
		Definition -1 mark	
		For significance -1 mark x 2=2 marks	
	iii	Discuss the Principle Component Analysis (PCA) technique with taking the example of 2D datasets?	5
		For defining PCA -2 marks	
		For 2-D datasets example -3 marks	
OR	iv	Apply the factor analysis techniques on Questionnaire data sets?	5
		For Questionnaire dataset -2 marks	
		For applying factor analysis -3 marks	

Q.5	i.	What is the difference between estimation and prediction?	2
		For each difference -1 mark x 2=2 marks	
	ii	Explain any model of performing statistical inference?	3
		For model definition -1 mark	
		For model explanation -2 marks	
	iii	Discuss the various steps involved for calculating the confidence interval with an example?	5
		For explaining each step operation -1 mark x 5=5 marks	
OR	iv	What do you mean by the term "Null Hypothesis"?	5
		Also discuss Type-I and Typ1-II error associated with it?	
		For definition -1 marks	
		For Type-I error explanation -2 marks	
		For Type-2 error explanation -2 marks	

Q.6	i.	Explain the importance of multivariate statistics?	2
		For each importance -1 mark x 2=2 marks	
	ii	Compare Two-Sample t-Test with Two-Sample z-Test?	3
		For each comparison -1 mark x 3=3 marks	
	iii	Discuss the various steps for performing Chi-square test for homogeneity?	5
		For explaining each step operation -1 mark x 5=5 marks	
OR	iv	Write down the steps for calculating the Two-Sample z-Test for two population proportions?	5
		For explaining each step operation -1 mark x 5=5 marks	