



**KONERU LAKSHMAIAH
EDUCATION FOUNDATION**
(Deemed to be University, Estd. u/s. 3 of UGC Act 1956)

B.Tech - Even Sem : End Semester Exam
Academic Year:2021-2022
19CS3275S - BIG DATA ANALYTICS
Set No: 4

Time:		Max.Marks: 100					
S.NO	Answer All Questions	Choice	Options	Marks	CO	CO BTL	COI BTL
1.	Define Big Data and recall some examples of Big Data. List various types of repositories and explain	choice Q-2		10Marks	CO1	2	1
2.	Explain the characteristics of Big Data. Illustrate various types of data structures of Big Data			10Marks	CO1	2	2
3.	Problem Statement: A clinical trial is run to compare weight loss programs and participants are randomly assigned to one of the comparison programs and are counseled on the details of the assigned program. Participants follow the assigned program for 8 weeks. The outcome of interest is weight loss, defined as the difference in weight measured at the start of the study (baseline) and weight measured at the end of the study (8 weeks), measured in pounds. Low Calorie(8,9,6,7,3) Low Fat(2,4,3,5,1) Low Carbohydrate(3,5,4,2,3) Control(2,2,-1,0,3) Is there a statistically significant difference in the mean weight loss among the four diets? Infer using the ANOVA five-step approach.	choice Q-4		15Marks	CO1	2	2
4.	Show the Data Analytics Life Cycle and explain the first three phases in detail.			15Marks	CO1	2	2
5.	Recall the concept of N-Fold cross-validation.	choice Q-6		10Marks	CO2	2	1
6.	Illustrate about MapReduce and State the need of YARN			10Marks	CO2	2	1
7.	Illustrate the concept of Topic Modeling. Illustrate the LDA Model. Highlight the concept of Sentiment Analysis.	choice Q-8		15Marks	CO2	2	2
8.	Illustrate the concept of Time Series Analysis. Explain about the components of Time Series Analysis. Differentiate Univariate and Multi-Variate Time Series Analysis.			15Marks	CO2	2	2
9.	Differentiate Pig, Hive and HBase	choice Q-10		10Marks	CO3	3	1
10.	Recall the importance of Combiner and a Mapper			10Marks	CO3	3	2
11.	Utilize the domain knowledge and develop one model and specify how to execute it. 1. Hive 2. HBase	choice Q-12		15Marks	CO2	2	2
12.	Model an example pig script and specify the commands to execute			15Marks	CO2	2	2
13.	State the four main deliverables of Data Science Projects	choice Q-14		10Marks	CO4	3	1
14.	Illustrate the four major categories of NoSQL tools			10Marks	CO4	3	1
15.	Develop sequence of steps to perform k-means using MADLib and SQL statements for required processing	choice Q-16		15Marks	CO4	3	1
16.	Illustrate the following with suitable examples 1. Joins 2. Inner Joins 3. Outer Joins 4. Left Outer Joins 5. Right Outer Joins.			15Marks	CO4	3	2

[object HTMLDivElement]