



B.Tech - Even Sem : End Semester Exam
Academic Year:2021-2022
19CS3275A - BIG DATA ANALYTICS
Set No: 2

Time:		Max.Marks: 100					
S.NO	Answer All Questions	Choice	Options	Marks	CO	CO BTL	COI BTL
1.	Compare Business Intelligence and Data Science.	choice Q-2		10Marks	CO1	2	1
2.	Illustrate seven key roles for successful analysis of the project..Illustrate various types of data structures of Big Data.			10Marks	CO1	2	2
3.	A global fast food chain wants to venture into a new city by setting up a store in a popular mall in the city. It has shortlisted 2 popular locations A and B for the same and wants to choose one of them based on the number of foot falls per week.The company has obtained the data for the same and wants to know which of the two locations to choose and if there is significant difference between the two.Here's how the footfalls per day (in 1000's) looks like for 10 randomly chosen days: Loc_A: (116.2,121.3,115.6,124.1,110.8,120.4,112.3,130.5,140.4,125.6) Loc_B: (110.0,110.0,104.3,120.1,109.4,100.2,110.1,107.4,131.8,125.1) $\bar{x}_{LocA}=121.72$ $\bar{x}_{LocB}=112.84$ $\bar{x}_{LocA} - \bar{x}_{LocB} = 8.88$ Location A gets 8.88k more footfalls on an average over a period of 10 days. Is an additional avg footfall of 8.88k measured over 10 days, enough to say that Loc A receives different footfalls than Loc B? Infer whether it is statistically significant using Welchs' Test. Given, t_{critic} is 2.101	choice Q-4		15Marks	CO1	2	2
4.	Illustrate six phases of Data Analysis Life Cycle			15Marks	CO1	2	2
5.	Show how to determine the no. of clusters.	choice Q-6		10Marks	CO2	2	1
6.	Compare the models of Linear Regression and Logistic Regression with an example model.			10Marks	CO2	2	2
7.	Illustrate the following 1. Bag of Words 2. Term Frequency 3. TFIDF	choice Q-8		15Marks	CO2	2	2
8.	Illustrate the need of confusion matrix and an example representation.			15Marks	CO2	2	2
9.	Experiment on hive for the following Create External table load the data manually	choice Q-10		10Marks	CO3	3	1
10.	Utilize the domain knowledge and develop one model and specify how to execute it. 1. Hive 2. HBase			10Marks	CO3	3	2
11.	Model an example pig script and specify the commands to execute	choice Q-12		15Marks	CO3	3	3
12.	Illustrate the architecture of HDFS. Classify various tools of Hadoop Eco System			15Marks	CO3	3	3
13.	Illustrate the following Set Operations with Suitable Examples Union Intersection Except	choice Q-14		10Marks	CO4	3	2
14.	Illustrate the four major categories of NoSQL tools			10Marks	CO4	3	1
15.	Model a user defined function to translate text strings for female (F) and male (M) to 0 and 1, respectively. Apply the designed user define function for any sql statement	choice Q-16		15Marks	CO4	3	3
16.	Summarize the key outputs for each of the main stakeholders of an analytics project and what they usually expect at the conclusion of a project			15Marks	CO4	3	3

[object HTMLDivElement]