DATA ANALYTICS

IMPORTANT QUESTIONS

YOUTUBE , TELEGRAM – SHORTNOTES4U

1.	What do you understand by regression modeling? Mention the three purpose of modelling?	Understand	CO2
2.	Differentiate between structured, semi-structured, and unstructured data?	Analyze	**CO1**
3.	What are the primary sources of data collection? Explain different types of source of data?	Understand	**CO1**
4.	Differentiate between classification and regression with examples?	Analyze	CO2
5.	Differentiate between data analytics vs data analysis?	Analyze	CO1

6.	What do you mean by data analytics? Explain the types of analytics?	Understand	CO1
7.	Explain the five characteristics of data? Describe four application of data analytics?	Understand	CO1
8.	Describe the basic difference between univariate, bivariate, and multivariate analysis with examples?	Understand	CO2
9.	What are the main three differences between data, information, knowledge, and wisdom? Explain different types of variables?	Understand	CO1

10.	(i).Explain the inference rules with examples? What do you understand by rule induction? [4] (ii).A man is known to speak the truth 2 out of 3 times. He throws a die and reports that the number obtained is a four. Find the probability that the number obtained is actually a four?...[3.5]	Understand	CO1
11.	Describe the architecture of Big Data? How is it different from data warehouse?	Understand	CO1
12.	Explain the following modelling with examples 1. Naive Bayes 2. Support Vector Machine. 3.logistic regression 4. Neural Network. 5. Linear system and Non Linear Dynamic in the context of time series data.	Understand	CO2
13	What are the five steps of analytics life cycle? What do you understand by confusion matrix and define at least three metrics of confusion matrix?	Understand	CO1

14.	What is Big Data? Explain characteristics of Big Data?	Understand	CO1
15.	Explain different phases of Data Analytics life cycle?	Understand	CO1
16.	How is Naïve Bayes different from Support Vector Machine?	Analyze	CO2
17.	Explain Hypothesis testing with example?	Understand	CO1
18.	What is regression? Explain any one type of regression in detail?	Understand	CO2
19.	What are the components of time series Analysis?	Understand	CO2
20	What is the definition of real time data?	Understand	CO3
21.	What do you understand by frequent data in DataStream?	Understand	CO3
22	Consider the following set of points: {(-2, -1), (1, 1), (3, 2)} a) Find the least square regression line for the given data points. b) Plot the given points and the regression line in the same rectangular system of axes.	Apply	CO2
23	What is fuzzy logic? Mention four application of Fuzzy logic?	Understand	CO2

24.	Explain real time data analytics engine in Big data to analysis the real time data. Explain the three Real time applications?	Understand	CO 3
25.	What is Sampling data in stream? What is Real Time Analytics Platform (RTAP)?	Understand.	CO3
26.	What do you understand by DataStream? Differentiate between Data Base Management System and? Data Science Management System?	Understand	CO3
27	Explain Non Linear SVM with suitable examples? How linear SVM is different from linear regression?	Understand	CO 2
28	What do you understand by stochastic search methods? Explain with one case studies in the terms of probability distribution?	Understand	CO 2

29.	What do you understand by Dimensional Reduction? Explain the working of PCA? How to Apply it? When to apply it?	Understand	CO2
30.	How can outlier detection be performed with linear regression? What is least square method? The relationship between the temperature and the number of passengers who take the bus has been studied by the city's transportation department. **S.No.** **Temp.** **Passengers**	Analyze	CO2

	1 42 173 2 37 149 3 46 185		
	4 30 123 5 50 201 6 43 174 7 43 175 8 46 188 9 46 186 10 49 198 For the case study given, 1. Find the least square reg line. 2. Predict the no of passen 45 3. Find the error ◦F		
31.	Differentiate between Fuzzy logic and Boolean logic? Illustrate the diagram of fuzzy decision tree? Apply the fuzzy logic to evaluate the degree of fastness of the car. Explain with suitable examples.	Understand	CO 2
32.	Illustrate the working of Blooms filter with the help of an example.	Apply	CO3
33.	Explain the working of Flajolet Martin Algorithm with an example	Understand	CO3
34.	Describe the concept of PCY & DGIM algorithms with example	Understand	CO4
35	What is boxplot analysis? Explain it with an example. Explain the descriptive statistics, and exploratory data analysis in R.	Understand	CO5
36	Explain about R, data types in R and list the various features of R.	Understand	CO5
37	What do you understand by NOSQL & S3 database?.What do you mean by sampling in data stream?	Understand	CO5
38	Expain the concept of counting of ones in a window?	Understand	CO3
39	What is basic difference between project clustering and k-mediods?	Analyze	CO4
40	What do you understand by clustering in non-euclidean space, clustering for streams and parallelism.	Understand	CO4

41	Discuss the library used in R for visualization? Explain various 2-Dimensional and 3-Dimensional graph and where can be utilized?	Understand	CO5

42.	Explain various types of clustering? Explain k-means & kmedoids clustering algorithm with use cases. Explain the concept of TYPE1& TYPE2 errors.	Understand	CO4
43	What do you mean by the apriori algorithm. How is it usefull in market basket analysis with an example.	Understand	CO3
44	Explain the various data analysis technique?.What is market basket analysis? What do you understand by handling large data sets in main memory. Explain any limited pass algorithm?	Understand	CO4
45	The transactional data for an all electronics branch is as follows, find the frequent itemset and generate association rules with confidence values. **Tid** | **List of Item_IDs** T100 | I1,I2,I5 T200 | I2,I4 T300 | I2,I3 T400 | I1,I2,I4 T500 | I1,I3 T600 | I2,I3 T700 | I1,I3 T800 | I1,I2,I3,I5 T900 | I1,I2,I3 We will consider minimum support level as 0.5. ie. 4. and minimum confidence 60%	Apply	CO4
46	Describe the following terms in more than 80 words for each with suitable diagram and examples: a.)Bigdata b).MapReduce c).HDFS d).Hive e).Pig f).HBASE g).Map R h).Sharding	Understand	CO5

47.	Illustrate the concept of clique and community using percolation method with example? How is it related to cloud computing? why is it different from k-means & k-mediods algorithm.	Apply	CO4
48	What do you understand by KNN and Hierarichal Clustering?.Use the data and group them using k-means clustering algorithm. Show calculations of centroid. Height Weight 185 72 170 56 168 60 179 68 182 72 188 77 180 71 180 70 183 84 180 88 180 67 177 76	Apply	CO3
49	Explian about rule for forming the bucket?.Here is a collection of twelve baskets. Each contains three of the six items 1 through 6. {1, 2, 3} {2, 3, 4} {3, 4, 5} {4, 5, 6} {1, 3, 5} {2, 4, 6} {1, 3, 4} {2, 4, 5} {3, 5, 6} {1, 2, 4} {2, 3, 5} {3, 4, 6} Suppose the support threshold is 4. On the first pass of the PCY Algorithm we use a hash table with 11 buckets, and the set {i, j} is hashed to bucket i × j mod 11. (a) By any method, compute the support for each item and each pair of items. (b) Which pairs hash to which buckets? (c) Which buckets are frequent? (d) Which pairs are counted on the second pass of the PCY Algorithm?	Apply	CO3
Q50. a).How can you load a .csv file in R?			
b).What are the different components of grammar of graphics?			
c).What is Rmarkdown? What is the use of it?			