Roll No. | | | | | | | | | | |

ANNA UNIVERSITY (UNIVERSITY DEPARTMENTS)

B.E. / B. Tech. END SEMESTER EXAMINATIONS, NOV/DEC 2024

INFORMATION TECHNOLOGY

IT7702 – Data Analytics

(Regulation 2015)

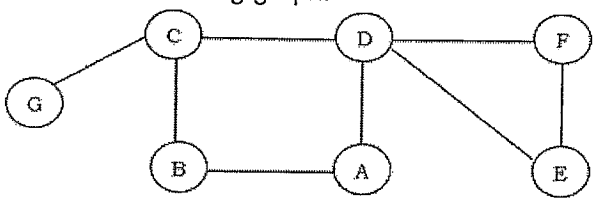
Time: 3hrs Max. Marks: 100

PART- A (10 x 2 = 20 Marks)
(Answer all Questions)

Q. No	Questions	Marks
1	What is big data? List its characteristics.	2
2	How reporting is different from analytics?	2
3	What is a kernel function? Mention any two kernel functions and its applications.	2
4	Give the differences between correlation and co-variance and also mention their usage in analytics.	2
5	What is a data node in Hadoop? Give its functions.	2
6	Mention any two important Map Reduce features that make it best for handling big data and state how.	2
7	Write the steps involved in the compilation of Hive query with an example.	2
8	Write the command(s) in R to read data from a data frame and display the column heading if available.	2
9	What is stream data? How it is different from traditional data?	2
10	Mention any four applications where stream data is used and state how it is handled in those applications.	2

PART- B (5 x 13 = 65 Marks)

Q. No	Questions	Marks
11 (a) (i)	Illustrate with necessary diagrams, features, pros and cons the different types of analytical sandbox configurations.	13
	OR	
11 (b) (i)	Identify, justify and explain apt sampling technique for the following scenarios: (I) A state's census data must be divided district wise and each district data must be treated as an independent population. (II) A small sized temperature data collected inside a car that is readily available and homogenous in nature. (III) Arranging the queries that hit a search engine based on time and then select every 10th query.	9

(ii)	State the null and alternative hypotheses for the following: (I) In a population of fish, approximately 42% are female. A test is conducted to see if, in fact, the proportion is less. (II) Twenty-nine percent of under-graduate students studying in a college get infected by flu each month.	4
12 (a) (i)	Consider the following graph:  Define and calculate the following for all the nodes: Eccentricity, Closeness and Betweenness.	9
(ii)	What is SSR, SSE and SST in regression analysis? Show mathematically the relationship between these metrics.	4
	OR	
12 (b) (i)	Calculate Eigen values and Eigen vectors for the matrix data given below. $A = \begin{matrix} 4 & 0 \\ 3 & -5 \end{matrix}$ Also explain how the derived Eigen vectors can be used for singular value decomposition.	13
13 (a) (i)	Draw the HDFS architecture and brief about the components, features and working of Hadoop with relevant example.	13
	OR	
13 (b) (i)	Explain how various relational algebraic operations can be represented as map and reduce functions.	13
14 (a) (i)	Compare and contrast the features of various types of NoSQL databases with relevant examples.	13
	OR	
14 (b) (i)	Explain different types of graph plots used for univariate and bivariate analysis with relevant R commands and examples.	13
15 (a) (i)	Discuss about the various sampling approaches used to perform stream data analytics with relevant real time examples. Also show mathematically how the sample sizes are determined.	13
	OR	
15 (b) (i)	State the rules of DGIM algorithm for counting the number of ones in a bit-stream. Consider the following stream: 10101100010111010010110. Divide the given stream into buckets as per DGIM rules and give the modified buckets after a new bit 1 arrives in the stream:	13

PART- C (1 x 15 = 15 Marks)

(Q.No.16 is compulsory)

<table>
<tr><th>Q. No</th><th>Questions</th><th>Marks</th></tr>
<tr><td>16.</td><td>Consider the following data where x is the time in years which an employee spend in a company and y is the hourly pay in rupees for 5 employees:
<table>
<tr><td>X (in years)</td><td>5</td><td>3</td><td>4</td><td>10</td></tr>
<tr><td>Y (in rupees)</td><td>250</td><td>200</td><td>210</td><td>350</td></tr>
</table>
Calculate the following:
(i) Mean, variance and standard deviation for X and Y
(ii) Covariance matrix
(iii) Apply linear regression modeling and give your inferences on the association between X and Y.</td><td>6
4
5</td></tr>
</table>

CONTROLLER OF EXAMINATIONS ADDITIONAL AU CHENNAI 600 025