No. of Printed Pages : 4

MCS-226

MASTER OF COMPUTER APPLICATIONS (MCA-NEW)

Term-End Examination

June, 2023

MSC-226 : DATA SCIENCE AND BIG DATA

Time : 3 Hours *Maximum Marks : 100*

Weightage : 70%

Note : *Question No. 1 is compulsory. Attempt any* ***three*** *questions from the rest.*

---

1. (a) Define Data Science. Give advantages of Data Science in an organization. 5

   (b) Explain Bayes' Theorem with suitable equation and example. 5

   (c) What is a Histogram ? How does Histogram differ from Bargraph ? Briefly discuss the utility of Histogram in Data Science. 5

   (d) What is Hadoop MapReduce ? Give its advantages. Also, discuss how <key-value> pair mechanism facilitates MapReduce programming. 5

P. T. O.

(e) In context of Data Science, what is Apache SPARK ? How does Apache SPARK differ from Hadoop ? 5

(f) What are Data Streams ? How do Data Streams differ from Databases ? Why mining of data streams is considered as a challenging process in Data Science ? 5

(g) Explain PageRank algorithm, with suitable example. 5

(h) What are Dataframes in 'R' programming ? Give characteristics of Dataframes. 5

2. (a) Write the syntax to create the following plots in 'R' : 5

(i) Bar charts

(ii) Box plots

(iii) Histogram

(iv) Line graphs

(v) Scatter plots

(b) Differentiate between Linear Regression and Multiple Regression, with suitable example for each. 5

(c) What are Decision Trees ? What are categorical variables and continuous variables ? How do these two variables relate to decision trees ? Explain the role of entropy and information gain in decision trees. 10

3. (a) Compare qualitative data with quantitative data. What do you understand by the term "Measurement scale of data" ? Give characteristics of measurement scales of data. List various measurement scales with suitable example for each. 10

(b) What is a Random Variable ? Differentiate between Discrete Random Variable and Continuous Random Variable. 5

(c) What is a Heat Map ? Give uses and best practices for Heat Maps. 5

4. (a) Explain the following operations of map-reduce with suitable example and supporting block diagram : 10

   (i) Splitting

   (ii) Mapping

   (iii) Shuffling

   (iv) Reducing

   (b) Explain the term Data lake. Briefly discuss the key capabilities of data lake. 5

   (c) Compare Land Mark Model and Sliding Windows Model for data stream processing. 5

5. Write short notes on the following : 4×5=20

   (a) Different mechanisms of finding PageRank

   (b) Chi-square test

   (c) Association Rules

   (d) Predictive Analysis

   (e) HIVE and its utility in Data Science