Roll No ....

CS-7005(2)-CBGS

B.E. VII Semester

Examination, June 2020

Choice Based Grading System (CBGS)

Data Science and Big Data

Time : Three Hours

Maximum Marks : 70

Note: i)   Attempt any five questions.

ii)   All questions carry equal marks.

1. a)   Give a detailed note on features and limitations of Python programming.   7

    b)   How analytical tools have evolved from graphical user interfaces to point solutions to data visualization tools?   7

2. a)   Discuss why is big data analytics important.   7

    b)   For each of the following scenarios, state which one is more likely to lead to the rejection of the null hypothesis

    i)   A one-tailed or two-tailed test

    ii)   .05 or .01 level of significance

    iii)   A sample size of $n = 144$ or $n = 444$

                                                                 7

3. A teacher draws a sample of 5, 12-year-old children from the school's population and records their heights as follows:

    {124, 124, 128, 130, 127}

    Assume that the heights have a normal distribution where both μ and σ are unknown. Calculate a two-tailed 95% confidence interval for the mean height of 12-year-olds.   14

4. Write short notes on any two: 14
   a) Logistic regression
   b) Back propagation algorithm
   c) Issues in machine learning

5. a) Explain about the Information Retrieval.   7
   b) Write the advantages and disadvantages for classic models which are used in IR and discriminate their techniques.   7

6. a) Briefly explain Web search architectures.   7
   b) Write the formal characterization of IR models.   7

7. a) Describe the working of Map reduce with an relevant example.   7
   b) Discuss the points to be considered while designing a file system in Map-reduce.   7

8. a) The product moment correlation coefficient between the random variables W and X is 0.71 and between the random variables Y and Z is –0.05. For each of these pairs of variables, sketch a scatter diagram which might represent the results which gave the correlation coefficients.   7
   b) The values of independent variable X and dependent value Y are given below: 7

X	Y
0	2
1	3
2	5
3	4
4	6

   Find the least square regression line y = ax + b. Estimate the value of y when x is 10.

******

CS-7005(2)-CBGS