SEAT No. :

PB145

[Total No. of Pages :2

[6269]- 359

T.E. (I.T.) (Insem)
DATA SCIENCE & BIG DATA ANALYTICS
(2019 Pattern) (Semester - II) (314452)

Time : 1 Hour]   *[Max. Marks : 30*
Instructions to the candidates:
  1) Answer Q.1 or Q.2, Q.3 or Q.4.
  2) Figures to the right indicate full marks.

Q1) a) "Data explosion plays an import role in Big Data" Justify the statement with proper explanation along with examples. **[8]**

b) Differentiate between Big Data and small Data with their processing architecture. **[7]**

OR

Q2) a) Define the relation between Artificial Intelligence, Statistical Learning and Machine Learning with sample example. **[8]**

b) Differentiate between Data ware house and Data mining with the Data Science perspective. **[7]**

Q3) a) In the Dark Ages, Harvard, Dartmouth and Yale admitted only male students. Assume that, at that time, 80 percent of the sons of Harvard men went to Harvard and the rest went to Yale, 40 percent of the sons of Yale men went to Yale; and the rest split evenly between Harvard and Dartmouth; and of the sons of Dartmouth men, 70 percent went to Dartmouth, 20 percent to Harvard, and 10 percent to Yale. **[8]**

  i) Find the probability that the grandson of a man from Harvard went to Harvard.

  ii) Modify the above by assuming that the son of a Harvard man always went to Harvard. Again, find the probability that the grandson of a man from Harvard went to Harvard.

b) Explain Blooms filter along with one application. Derive the equation for probability distibution in Blooms filter. **[7]**

OR

P.T.O.

Q4) a) Explain the Flajolet Mart in Distance Sampling. Find the distinct Element from the element stream S=1,3,2,1,2,3,4,3,1,2,3,1
h (x) = (6x+1) mod 5. **[8]**

b) Only 1 in 1000 people have a rare disease A with TP=0.99 and FP=0.02 If one randomly tested individual is positive, what is the probability that they have a disease.

Given: P (A)=0.001, P(A_0)=0.000, P(B/A)=0.99, P(B/A_0)=0.02. **[7]**

 G3   G3   G3