Total No. of Questions : 4]

SEAT No. :

PA-10062

[Total No. of Pages : 1

[6009]-353

T.E.(Information Technology) (Insem)
DATA SCIENCE AND BIG DATA ANALYTICS
(2019 Pattern) (Semester-II) (314452)

Time : 1 Hour] *[Max. Marks : 30*

Instructions to the candidates:

1) *All questions are compulsory.*
2) *Figures to the right indicate full marks.*
3) *Assume suitable data if necessary.*
4) *Attempt Q.1 or Q.2, Q3 or Q4.*

Q1) a) Explain 6V's for defining Big Data along with the factors responsible for data explosion? **[8]**

b) List and explain data processing infrastructure challenges in Big Data with suitable example. **[7]**

OR

Q2) a) List and explain choices for reengineering the data warehouse? **[8]**

b) Explain shared-everything and shared nothing architectures in detail with respect to Big data? **[7]**

Q3) a) Explain the following terms. **[7]**

i) Expectation

ii) Pair wise independence.

b) Given that a person last purchase was coke. there is a 90% chance that his next purchase will also be coke. If a person's last purchase was Pepsi, there is an 80% chance that his next purchases will also Pepsi. **[8]**

i) Given that a person is currently a Pepsi purchaser, what is the probability that he will purchase Coke two purchases from now?

ii) Give that a person is currently a Coke drinker, what is the probability that he will purchase Pepsi three purchases from now?

OR

Q4) a) Explain Flajolet Martin Distance Sampling? Find the distinct element from the element stream 4,2,5,9,1,6,3,7. Consider the Hash function h(x) = (3x+7) mod 32. **[8]**

b) Find the variance and standard deviation for the following data set: 70, 60, 72, 42, 86 **[7]**

● ● ●