Total No. of Questions : 6]

SEAT No. :

P53

[Total No. of Pages : 2

TE/INSEM/APR-58
T.E. (IT) (Semester - II)
314454 : DATA SCIENCE AND BIG DATA ANALYTICS
(2015 Pattern)

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

Instructions to the candidates:

1) ***Attempt Q. 1 or Q. 2, Q. 3 or Q. 4 and Q. 5 or Q. 6.***
2) ***Neat diagrams must be drawn wherever necessary.***
3) ***Figures to the right side indicate full marks.***
4) ***Assume suitable data if neeessary.***

Q1) a) What are ihe pitfalls of data warehouse? Why companies are shifting to Big Data using Hadoop. **[4]**

b) Draw and explain Big Data Processing Architecture with technologies used at each of the stage of big data processing. **[6]**

OR

Q2) a) Explain Different Learning Approaches in Big Data. Explain with example. **[6]**

b) Explain Data Science and its various applications. **[4]**

Q3) a) What is the role of different kinds of distributions in Big Data? Explain it with example. **[4]**

b) Only 1 in 1000 people has rare disease. Given True Positive = 0.9 and False positive = 0.02. If randomly tested individual is positive. What is the probability that they have a disease. **[6]**

OR

P.T.O.

Q4) a) A petrol station owner is considering the effect on his business (Superpet) of a new petrol station (Global) which has opened just down the road. Currently (of the total market shared between Superpet and Global) Superpet has 80% of the market and Global has 20%. Analysis over the last week has indicated the following probabilities for customers switching the station they stop at each week: **[6]**

		To	
		Superpet	Global
From	Superpet	0.75	0.25
	Global	0.55	0.45

- What will be the expected market share for Superpet and Global after another two weeks have past?
- What would be the long-run prediction for the expected market share for Superpet and Global?

b) Explain Blooms filter with example. State the applications of Blooms filter in Big Data **[4]**

Q5) a) Explain the-GFS system with respect to (i) Architecture (ii) Types of metadata (iii) Strengths. **[6]**

b) Explain the role of Job and Task tracker in the execution and processing of job within Hadoop environment, with architecture. **[4]**

OR

Q6) a) Draw and explain Map Reduce architecture. **[6]**

b) Write a short note on Textual ETL processing. **[4]**