

Total No. of Questions : 10]

SEAT No. :

P1770

[Total No. of Pages : 3

[5460] - 600

**T.E. (Information Technology)
Data Science and Big Data Analytics
(2015 Pattern) (Semester - II)**

Time : 2½ Hours]

[Max. Marks :70

Instructions to the candidates:

- 1) *Answer Q.1 or Q.2, Q.3 or Q.4, Q.5 or Q.6, Q.7 or Q.8, Q.9 or Q.10.*
- 2) *Draw neat diagrams whenever necessary.*
- 3) *Figures to the right indicate full marks.*
- 4) *Assume suitable data if necessary.*

- Q1)** a) State one example of big data and explain how all V's are applied for big data 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

- Q2)** a) How Big Data Analytics is useful for improving agriculture and life of farmers. **[4]**
- b) A computer device can be either in a busy mode (state 1) processing a task, or in an idle mode (state 2), when there are no tasks to process. Being in a busy mode, it can finish a task and enter an idle mode any minute with the probability 0.2. Thus, with the probability 0.8 it stays another minute in a busy mode. Being in an idle mode, it receives a new task any minute with the probability 0.1 and enters a busy mode. Thus, it stays another minute in an idle mode with the probability 0.9. The initial state is idle. Let X_n be the state of the device after n minutes. Find the distribution of X_2 . **[6]**

P.T.O.

Q3) a) What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with proper diagram. **[5]**

b) What is Textual ETL processing? Explain different components of textual ETL processing. **[5]**

OR

Q4) a) Explain the concept of Blocks and Heartbeat mechanism in HDFS Architecture. **[5]**

b) List and explain any five hadoop shell commands with syntax. **[5]**

Q5) a) What is need of integrating R and Hadoop? Explain the process of R and Hadoop Integration. **[9]**

b) How Missing values and categorical variables are preprocessed before building model? Explain with example. **[9]**

OR

Q6) a) Explain different types of analysis in detail with example. **[9]**

b) Explain different techniques of data visualization in detail. **[9]**

Q7) a) What is the need of data visualization? Also explain advantages of data visualization. **[8]**

b) Draw and Explain detail architecture of Text Mining and explain why it is required. **[8]**

OR

Q8) a) What do you mean by Model Planning and Model building? Explain tools for model building. **[8]**

b) Explain various Challenges in Big data visualization and explain the mechanism to overcome the challenges. **[8]**

- Q9)** a) What is Mobile Analytics? What is the importance of Mobile analytics?[8]
b) Explain the process of social media data analytics with example. [8]

OR

- Q10)** a) What are various Responsibilities of Data Scientist? Explain with Diagram. [8]
b) Write note on [8]
i) Tag Cloud
ii) Cluster-gram
iii) Motion Chart
iv) Google Chart API

