

Total No. of Questions : 10]

SEAT No. :

**P3975**

**[5561]-679**

[Total No. of Pages : 2

**B.E.(Computer Engineering)  
DATA ANALYTICS  
(2015 Pattern) (Semester - I) (410243)**

*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) Neat diagrams must be drawn wherever necessary.
- 3) Figures to the right side indicate full marks.
- 4) Assume suitable data if necessary.

- Q1)** a) What is big data? Explain 3V's of Big Data. [5]  
b) Draw Data Analytics Lifecycle & give brief description about all phases. [5]

OR

- Q2)** a) Write a case study on Global Innovation Network & Analysis (GINA). [5]  
b) Explain Null Hypothesis & Alternative Hypothesis. [5]

- Q3)** a) How Wilcoxon Rank-Sum Test works? [5]  
b) Explain Type 1 and Type 2 errors. [5]

OR

- Q4)** a) Write an Apriori Algorithm. [5]  
b) Define following terms with example : Confidence and Lift. [5]

- Q5)** a) Explain following Decision Tree Algorithms : [9]  
i) ID3 Algorithm  
ii) C4.5  
iii) CART  
b) How Naive Baye's classification works? Give its applications. [8]

OR

**P.T.O.**

- Q6) a)** Explain following terms : **[9]**
- i) Bagging
  - ii) Boosting
  - iii) Random forest

- b) What is data visualization? Describe any four data visualization techniques. **[8]**

- Q7) a)** Why is it difficult to visualize Big Data? Also explain analytical techniques used in Big Data Visualization. **[9]**

- b) Explain various tools to visualize Big Data. (Any four) **[8]**

OR

- Q8) a)** What is Map-Reduce? Explain working of Map-Reduce with example. **[9]**

- b) Explain HDFS with respect to NameNode, DataNodes, Secondary NameNode with example. **[8]**

- Q9) a)** Explain following terms : **[8]**

- i) Smoothing
- ii) Confusion matrix

- b) Explain Data Visualization Tool - Tableau. **[8]**

OR

- Q10) a)** Explain following terms : **[8]**

- i) Key-value store
- ii) Document store
- iii) Column family store
- iv) Graph Databases

- b) Why communication is important in data analytics lifecycle projects? **[8]**

