

Total No. of Questions : 10]

SEAT No. :

P3424

[Total No. of Pages : 2

[5670]-700

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, Q3 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) Draw and Explain Big Data Ecosystem. [5]

b) Explain current analytical architecture with diagram. [5]

OR

Q2) a) What is mean by Data conditioning and data visualization? Enlist tools used during data preparation phase. [5]

b) Write a case study: Global Innovation Network and Analysis. [5]

Q3) a) How k-means algorithm works? [5]

b) Explain Wilcoxon Rank-Sum Test. [5]

OR

Q4) a) How association rules are helpful in developing business strategy? [5]

b) How to improve 'Apriori's efficiency? [5]

Q5) a) Explain following decision the algorithms: [9]

i) ID 3

ii) C4.5

iii) CART

b) What is decision tree? Explain various terms used in Decision Tree. [8]

OR

P.T.O.

- Q6)** a) Explain “Bagging”, “Boosting”, and “Random Forest”. [9]  
b) Explain various tools to visualize Big Data (Any four) [8]

- Q7)** a) Explain data visualization with respect to 1-D, 2-D and 3-D data. [9]  
b) Why it is difficult to visualize big data? [8]

OR

- Q8)** a) Explain working of Apache Hadoop with HDFS and Map Reduce. [9]  
b) Explain following terms: [8]  
i) Pig ii) Hive  
iii) HBase iv) Mahout

- Q9)** a) Explain various use-cases of mapreduce for unstructured data. [8]  
b) Explain data visualization Tool-Tableau [8]

OR

- Q10)** a) Explain following NOSQL databases: [8]  
i) Key value store  
ii) Document store  
iii) Column family store  
iv) Graph database  
b) What are the key outputs from each stake holder (any four) in analytical project? [8]

