Total No. of Questions : 8]

SEAT No. :

P3333

[Total No. of Pages : 2

[5461]-593
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.***
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) Explain Data Analytic Life cycle. **[8]**

b) When do we use Wilcoxon rank-sum test? Write steps in the test. **[6]**

c) Explain linear regression with example. **[6]**

OR

Q2) a) Compare BI Vs.Data science. **[6]**

b) Explain k-means clustering algorithm. What are its drawbacks? **[7]**

c) Explain Apriori association rule mining algorithm. **[7]**

Q3) a) Explain Bayes 'theorem.Explain Naive Bayes' classifier. **[8]**

b) Explain any three of classification performance measures. **[6]**

c) What is classification? List the different classifiers. **[3]**

OR

Q4) a) What is decision tree? Explain how decision tree is constructed using ID3 algorithm. **[8]**

b) Explain the following: **[3]**

i) Conditional probability.

ii) Posterior probability.

c) Explain the following with their significance: **[6]**

i) Entropy

ii) Information gain

iii) Gain ratio

P.T.O.

Q5) a) What is data visualization?Explain any four data visualization Techniques **[9]**

b) What are the challenges in Big data visualization? **[8]**

OR

Q6) a) Explain how data is visualization is done or visually represented, if data is 1-D, if data 2-D and data is 3-Diamentional? **[6]**

b) Explain Big data visualization tools in short (any four tools). **[8]**

c) Explain analytical techniques used in Big data visualization. **[3]**

Q7) a) Explain use cases for analytics for unstructured data. **[5]**

b) Explain MapReduce paradigm with example. **[6]**

c) Explain Hadoop Distributed File System. **[5]**

OR

Q8) a) Explain the Hadoop Ecosystem in detail with Pig, Hive, HBase and Mahout. **[8]**

b) Give a brief review of the key outputs for each of the main any four stakeholders of an analytics project. **[4]**

c) What are four major categories of NOSQL Tools (stores)? **[4]**