Total No. of Questions : 8]

SEAT No. :

P-3153

[Total No. of Pages : 2

[6003]-354

T.E. (Computer Engineering)
Data Science and Big Data Analytics
(2019 Pattern) (Semester - II) (310251)

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.

2) Neat diagram must be drawn whenever necessary.

3) Figures to the right indicate full marks.

4) Assume suitable data if necessary.

5) Use of Scientific Calculalor is permitted.

Q1) a) What is Model Building elaborate this phase of data analytics with the help of suitable example. **[9]**

b) Explain any three sources of Big Data. Differentiate BI versus Data science. **[8]**

OR

Q2) a) What are the three characteristic of Big Data and what are the main consideration in processing Big Data. **[8]**

b) Explain Descriptive, Diagnostic, Predictive analytics. **[9]**

Q3) a) Explain why decision tree are used. Draw a sample decision tree and explain its parts. **[9]**

b) How Apriori Algorithm works, explain with suitable example? **[9]**

OR

Q4) a) What is data preprocessing? Explain in details about handling missing data and transformation of data. **[9]**

b) Explain Naïve Bayes' classifier and it applications. **[9]**

P.T.O.

Q5) a) What is text processing? Explain TF-IDF with example. **[8]**

b) With suitable example ,explain the steps involved in k-means algorithm. **[9]**

OR

Q6) a) Define following terms with respect to confusion matrix : **[8]**

    i) Accuracy

    ii) Precision

    iii) Recall

    iv) AUC-ROC

b) Explain k-fold Cross Validation & Random Subsampling. **[9]**

Q7) a) With a suitable example, draw a Histogram, boxplot and explain its usages. **[9]**

b) Describe the data visualization tool Tableau. List of data visualization tools. **[9]**

OR

Q8) a) What is Data Visualization? Describe the challenges of data visualization. **[9]**

b) Explain architecture of Apache-Pig. **[9]**

ఇఇఇ