[5870] - 1133

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, Q.3 or Q.4, Q.5 or Q.6, Q.7 or Q.8.*
1) *Neat diagrams must be drawn whenver necessary.*
2) *Figures to the right side indicate full marks.*
3) *Use of logarithmic tables slide rule, mollier charts, electronic pocket calculator and steam tables is allowed.*
4) *Assume suistable data, if necessary.*

Q1) a) What is driving data deluge? Explain with one example. **[9]**

  b) What is data science? Differentiate between Business Intelligence and Data Science. **[9]**

OR

Q2) a) What are the sources of Big Data. Explain model building phase with example. **[9]**

  b) Explain big data analytics architecture with diagram. What is data discovery phase. Explain with example. **[9]**

Q3) a) Explain various data pre-processing steps. Discuss essential python libraries for preprocessing. **[8]**

  b) What are association rules? Explain Apriori Algorithm in brief. **[9]**

OR

Q4) a) Explain the following

  i) Linear Regression

  ii) Logistic Regression **[8]**

  b) Explain scikit-learn library for matplotlib with example. **[9]**

P.T.O.

Q5) a) Write short note on

    i) Time series Analysis

    ii) TF - IDF. **[9]**

b) What is clustering? With suitable example explain the steps involved in k - means algorithm. **[9]**

OR

Q6) a) Write short note on

    i) Confusion matrix

    ii) AUC - ROC curve **[9]**

b) Discuss Holdout method and Random Sub Sampling methods. **[9]**

Q7) a) With a suitable example explain Histogram and explain its usages. **[8]**

b) Describe the Data visualization tool "Tableau". Explain its applications in brief. **[9]**

OR

Q8) a) With a suitable example explain and draw a Box plot and explain its usages. **[8]**

b) Describe the challenges of data visualization. Draw box plot and explain its usages. **[9]**

CB CB CB