PC1726

[6353]-43

T.E. (Computer Engineering)
DATA SCIENCE AND BIG DATA ANALYTICS
(2019 Pattern) (Semester-II) (310251)

Time : 2 ½Hours]   *[Max. Marks : 70*

Instructions to the candidates:

1) *Attempt Q.1 or Q.2, Q.3 or Q.4, Q.5 or Q.6, Q.7 or Q.8.*
2) *Figures to the right indicate full marks*
3) *Neat Diagram must be drawn wherever necessary.*
4) *Assume suitable data if necessary.*

Q1) a) Draw data analytics life cycle diagram and briefly explain its phases. **[6]**

b) Explain the various key roles for a successful analytics project. **[6]**

c) What are various sources of Big data. **[6]**

OR

Q2) a) Describe few applications of Big Data Analytics. **[6]**

b) Explain Data preparation phase of data analytics lifecycle. **[6]**

c) List common tools used for model building phase of data analytic. **[6]**

Q3) a) Define and explain Entropy and Information gain. Calculate the entropy of the following distribution **[9]**

Fruit Color	Taste	Count
Yellow	Sweet	10
Red	Sweet	5
Green	sour	15
Orange	sour	5

b) Explain Naïve Bayes Classifier. **[8]**

OR

P.T.O.

Q4) a)   Explain Apriori algorithm with suitable example. **[9]**

b)   Describe different categories of analytics **[8]**

Q5) a)   What is Hierarchical clustering? Explain hierarchical clustering algorithms. **[9]**

b)   Write a note on: **[8]**

i)   Holdout method

ii)   k-Fold Cross-Validation

OR

Q6) a)   What is Text analysis? Explain the different steps involved in the text analysis. **[9]**

b)   Write a note on Social network analysis. What are the applications of Social network analysis? **[8]**

Q7) a)   Explain Hive architecture with suitable diagram. Describe characteristics and features of hive. **[9]**

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

OR

Q8) a)   What is data visualization and objectives of data visualization? Why it is difficult visualize Big Data? **[9]**

b)   Write a note on Microsoft Power BI and Qlik **[9]**

❖❖❖❖