**[6262]-56**

## T.E. (Artificial Intelligence and Data Science)

## DATA SCIENCE

## (2019 Pattern) (Semester - II) (317529)

*Time : 2½ Hours]* *[Max. Marks : 70*

*Instructions to the candidates:*

1) *Solve questions 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 indicate full marks.*
4) *Assume suitable data if necessary.*

*Q1)* a) What is Data Analytics Life Cycle? **[6]**

b) Draw the diagram of the data analytics life cycle in big data and briefly explain the Model Planning phase. **[6]**

c) Write a note on the Data Preparation phase with its steps. **[6]**

**OR**

*Q2)* a) Explain in detail how the model-building phase is built by a team in the data analytics life cycle. **[6]**

b) List and explain the steps in the Data Preparation phase of the data analytics lifecycle. **[6]**

c) Explain challenges in the Model building phase. **[6]**

*Q3)* a) What do you mean by Linear Regression? Elaborate the types. **[6]**

b) Explain the Apriori algorithm with an example. **[6]**

c) Write a short note on the following: **[6]**

i) FP growth

ii) Decision Tree Classification

**OR**

*Q4)* a) Explain Data transformation using function and mapping. **[6]**

b) Write a short note on the following: **[6]**

i) Removing duplicates from the data set.

ii) Handling missing data.

c) Explain association rules with examples. **[6]**

*P.T.O.*

*Q5)* a) What is clustering? Explain hierarchical clustering with an example. **[6]**

b) Explain the Holdout method and Random Sub Sampling method. **[6]**

c) Discuss parameter tuning and optimization. **[5]**

OR

*Q6)* a) Short Note on: **[6]**

i) AUC-ROC Curves

ii) Elbow plot

b) What is clustering? Explain k - means algorithm. **[6]**

c) Write a short note on : **[5]**

i) Time series Analysis

ii) TF-IDF.

*Q7)* a) What is data Visualization? List and explain any one type of data visualization. **[6]**

b) With a suitable example explain and draw a Box plot and explain its usage. **[6]**

c) Discuss various challenges to Big data Visualization. **[5]**

OR

*Q8)* a) Define data visualization. What are the different methods of data visualization explain in detail. **[6]**

b) Explain in detail the Hadoop Ecosystem with a suitable diagram. **[6]**

c) With a suitable example explain the Scatter plot and explain its usage.**[5]**

✗ ✗ ✗