

Total No. of Questions : 8]

**PC1739**

SEAT No. :

[6353]-56

[Total No. of Pages : 2

**T.E. (Artificial Intelligence & Data Science)  
DATA SCIENCE  
(2019 Pattern) (Semester-II) (317529)**

*Time : 2½ Hours]*

*[Max. Marks : 70*

*Instructions to the candidates:*

- 1) *Solve 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) Explain in Detail Discovery Phase of Data Analytics lifecycle. [7]  
b) Explain in Detail Data Preparation Phase of Data Analytics lifecycle. [4]  
c) Explain in Detail Model Building Phase of Data Analytics lifecycle. [7]

OR

**Q2)** a) Explain in Detail Communication Result Phase of Data Analytics lifecycle. [8]  
b) Write a Short Notes on Global Innovation Social Network and Analysis (GINA). [10]

**Q3)** a) What are different python libraries used for data analysis? [3]  
b) Explain data preprocessing in detail. [10]  
c) Differentiate between Predictive, Descriptive, and Prescriptive data analysis. [4]

OR

**Q4)** a) What are association rules? Explain Apriori Algorithm in brief. [5]  
b) Write python program to apply the logistic regression model and check for goodness of fit. [7]  
c) What is the decisiontree? Explain with an example. [5]

**Q5)** a) What do you mean by text analysis? Why text analysis needs to be done? Explain the following text analysis steps with suitable example. [12]

- i) Part of speech (POS) tagging
- ii) Lemmatization
- iii) Stemming

b) Write short note on: [6]

- i) Time series Analysis
- ii) TF - IDF.

OR

**Q6)** a) What is clustering? With suitable example explain the steps involved in k - means algorithm. [8]

b) Discuss Holdout method and Random Sampling methods. [6]

c) Write short note on: [4]

- i) Confusion matrix
- ii) AUC - ROC curve

**Q7)** a) Write a short note on the following [6]

- i) Map reduce
- ii) Pig
- iii) Hive

b) Describe the Data visualization tool “Tableau”. Explain its applications in brief. [6]

c) With a suitable example explain Histogram and explain its usages. [5]

OR

**Q8)** a) Explain in detail the Hadoop Ecosystem with suitable diagram. [11]

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

