

Total No. of Questions : 8]

SEAT No. :

PD4260

[Total No. of Pages : 2

[6403]-56

T.E. (Artificial Intelligence and Data Science)

DATA SCIENCE

(2019 Pattern) (Semester - VI) (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) Explain phases of data analytic life cycle in detail. Draw a suitable diagram. [9]
- b) What is model building elaborate this phase of Data analytics with the help of suitable example. [8]

OR

- Q2)** a) Write a short note on : [9]
- i) ETL
 - ii) Common tools for model building
 - iii) Model selection for Data analytics
- b) List and explain the various activities involved in identifying potential data resources as a part of discovery phase in Data analytics life cycle. [8]

- Q3)** a) Write short note on the following : [9]
- i) Removing duplicates from data set.
 - ii) Data transformation
 - iii) FP Growth
- b) What are association rules? Explain Apriori Algorithm in brief. [8]

OR

P.T.O.

- Q4)** a) Explain Predictive, Descriptive and Prescriptive analytics with example.[9]
 b) What is Classification? Explain Decision Tree Classification with example.[8]

- Q5)** a) Write short note on the following : [9]
 i) Confusion matrix.
 ii) AUC-ROC Curves.
 iii) Elbow plot.
 b) Explain the following Text Analysis steps with suitable example : [9]
 i) Part-of-speech(POS)tagging
 ii) Lemmatization

OR

- Q6)** a) What is clustering? With suitable example explain the steps involved in Hierarchical Clustering. [9]
 b) What is confusion matrix? Given the confusion matrix, Calculate Accuracy, Precision, Recall, Error rate with description on diabetic risk[9]

Actual Classes	Predicted Classes		
	Classes	Diabetic Risk- Yes	Diabetic Risk- No
	Diabetic Risk- Yes	90	210
	Diabetic Risk- No	140	9560

- Q7)** a) List the challenges of Big Data Visualization. Explain the types of visualization with example. [9]
 b) Explain in detail the Hadoop Ecosystem with suitable diagram and different components. [9]

OR

- Q8)** a) Write a short note on the following : [9]
 i) Map reduce.
 ii) Pig
 iii) Hive
 b) With suitable example draw histogram and boxplot. Also explain its uses.[9]

