

Please check that this question paper contains 09 questions and 01 printed pages within first ten minutes.

[Total No. of Questions: 09]

[Total No. of Pages: 01]

Uni. Roll No. ....

Program: **B. Tech. (Batch 2018 onward)**

Semester: **8 th**

Name of Subject: **Data Science**

Subject Code: **PECS-119**

Paper ID: **17576**

Time Allowed: **03 Hours**

Max. Marks: **60**

**NOTE:**

- 1) Parts A and B are compulsory
- 2) Part-C has Two Questions Q8 and Q9. Both are compulsory, but with internal choice
- 3) Any missing data may be assumed appropriately

**Part – A**

**[Marks: 02 each]**

Q1.

- a) Differentiate between Data Science, Machine Learning and AI.
- b) Explain backward selection procedure in brief.
- c) Illustrate the process of handling the missing values and outliers.
- d) List out the differences between Data Science and Business Intelligence.
- e) Explain importance of data representation.
- f) Contrast Quantitative and Qualitative variables with example.

**Part – B**

**[Marks: 04 each]**

Q2. Write about:

- a) Data cleansing and transformation.
  - b) Dependent and Independent Variables.
- Q3. Explain Big Data Growth issues and challenges in detail.
- Q4. Distinguish concepts of under fitting and over fitting with example.
- Q5. Describe the normal distribution curve with its characteristics and deviation.
- Q6. Explain various techniques of data visualizations in detail.
- Q7. Distinguish b/w: a) NoSQL and SQL database  
b) Accuracy and Precision

**Part – C**

**[Marks: 12 each]**

Q8. Introduce and write the purpose of EDA (Exploratory Data Analysis) in detail.

**OR**

Calculate correlation coefficient for the following data by short-cut method:

<b>X</b>	10	12	14	18	20
<b>Y</b>	5	6	7	10	12

Q9. Describe Data Science Process with suitable flow diagram.

**OR**

Illustrate following:

- a) Structured and Unstructured data.
- b) Confusion matrix.
- c) Live data streams.

\*\*\*\*\*