Total No. of Questions : 8]

SEAT No. :

P-7626

[Total No. of Pages : 2

[6180]-146

T.E. (Information Technology)

DATA SCIENCE AND BIG DATA ANALYTICS

(2019 Pattern) (Semester - II) (314452)

Time : 2½ Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***

Q1) a) Explain the process of reading and writing a file in HDFS with neat diagram. **[8]**

b) List and explain any four Hadoop shell commands with syntax. **[4]**

c) Differentiate between SQL and NoSQL databases with example.What is the need to develop big data applications using NoSQL databases? **[6]**

OR

Q2) a) What is the need of map reduce in Big Data? Explain the stages involved in map reduce task with a suitable example? **[9]**

b) Explain Hadoop ecosystem in detail. **[9]**

Q3) a) What is data wrangling? Why do you need it? Explain data wrangling methods? **[9]**

b) What is categorical variable? Why do you need categorical variable encoding? With an example, explain one-hot encoding. **[9]**

OR

Q4) a) Draw and explain Architecture of HIVE. **[8]**

b) How missing values and categorical variables are preprocesses before building model? Explain with suitable example. **[4]**

c) Explain z-score normalization. For the following dataset carry out z-score normalization (standardization), X = 23, 29, 52, 31, 45, 19, 18, 27. **[6]**

P.T.O.

Q5) a) What is Data Visualization? What are the major challenges in big data visualization and how to overcome these challenges? **[6]**

b) Explain various techniques for visual data representation. **[6]**

c) Explain the following data visualization techniques. **[5]**

i) Candela

ii) D3.js

OR

Q6) a) Explain data visualization with respect to l-D, 2-D, 3-D data. **[6]**

b) Explain various analytical techniques used in big data visualization. **[6]**

c) Draw histogram with a suitable example and explain its usage. **[5]**

Q7) a) What is Porters valuation creation model? Explain porter's value chain analysis. **[9]**

b) What is social media analytics? Explain the process of social media data analytics. **[8]**

OR

Q8) a) What is text mining? Draw and explain text mining architecture and its use. **[8]**

b) Explain primary activities of Michal Porters value chain. **[5]**

c) How mobile analytics is different from social media analytics? **[4]**

□□□