Roll No. ⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚⬚   **Total No. of Pages : 02**

Total No. of Questions : 8

M.Tech(Computer Science & Engineering)E-II (2018 Batch) (Sem.–1)
DATA SCIENCE
Subject Code : MTCS-108-18
Paper ID : [75158]

Time : 3 Hrs.   **Max. Marks : 60**

INSTRUCTIONS TO CANDIDATES :
1.Attempt any FIVE questions out of EIGHT questions.
2.Each question carries TWELVE marks.

Q.1. (a) What is the difference between Data Analysis and Data Reporting? Explain with an example.

(b) Which are the various Data Science toolkit available and how they can be used? Explain.

Q.2. An advertisement for a tennis magazine states, *"If I'm not playing tennis, I'm watching tennis. And if I'm not watching tennis, I'm reading about tennis."* We can assume that the speaker can do at most one of these activities at a time. What is the speaker doing? Give reasons for your answer.

(a)   Playing tennis.

(b)   Watching tennis.

(c)   Reading about tennis.

(d)   None of the above.

Q.3. (a) Differentiate between Univariate, Bivariate and Multivariate analysis.

(b) During analysis, how do you treat Missing Values? Explain with an example.

Q.4. What do you understand by Recall and Precision? Explain with an example. Also, what we want a high Recall or a high Precision? How F1-measure is related to Recall and Precision and how it can be calculated?

Q.5.   Write a short note on how following machine learning technique can help in data analysis:

      (a) SVM

      (b) Linear Regression

      (c) Naive Bayes

Q.6.   (a) Discuss any five real-time applications of Data Science.

      (b) What are the various Data Visualization techniques available? Explain.

Q.7.   (a) Which language is more suitable for Text Analytics? R or Python? Why?

      (b) How do Data Scientists use Statistics? Explain with a real-world example.

Q.8.   (a) Why data cleansing is necessary for data analysis? Explain.

      (b) Can you explain the difference between a Test Set and a Validation Set?