

Roll No.

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Total No. of Pages : 02

Total No. of Questions : 17

MBA (2018 Batch) (Sem.-3)
DATA SCIENCES USING R
Subject Code : MBA-962-18
M.Code : 77058

Time : 3 Hrs.

Max. Marks : 60

INSTRUCTIONS TO CANDIDATES :

1. SECTION-A contains EIGHT questions carrying TWO marks each and students has to attempt ALL questions.
2. SECTION-B consists of FOUR Subsections : Units-I, II, III & IV. Each Subsection contains TWO questions each carrying EIGHT marks each and student has to attempt any ONE question from each Subsection.
3. SECTION-C is COMPULSORY and consists of ONE question carrying TWELVE marks.

SECTION-A

Explain briefly :

1. Data Pre-processing
2. Standard Deviation
3. Variance
4. Correlation
5. Machine Learning
6. Clustering
7. KDD
8. Validation

SECTION-B

UNIT-I

9. What are different components of Data Science and their roles?
10. Explain mean, median and mode using R software.

UNIT-II

11. What is SVM? Explain its use and applications.
12. Describe the concept of Bays theorem.

UNIT-III

13. Explain the process of clustering with k-means algorithm.
14. What is classification? How random forest is used?

UNIT-IV

15. How evaluation and validation is achieved using Bootstrap method?
16. Write a note on scalable machine learning.

SECTION-C

17. A librarian purchases 500 books for the library. How data mining techniques (clustering and classification) will be used to arrange the books into racks?

NOTE : Disclosure of Identity by writing Mobile No. or Making of passing request on any page of Answer Sheet will lead to UMC against the Student.