

DR. BABASAHEB AMBEDKAR TECHNOLOGICAL UNIVERSITY,
LONERE – RAIGAD -402 103
Winter Semester Examination – Dec. - 2019

Branch: Computer Science & Engineering M.Tech.

Subject:- Data Science (MTCE1201)

Date:-11/12/2019

Sem.:- II

Marks: 60

Time:- 3 Hr.

Instructions to the Students

1. Each question carries 12 marks.
2. Attempt **any five** questions of the following.
3. Illustrate your answers with neat sketches, diagram etc., wherever necessary.
4. If some part or parameter is noticed to be missing, you may appropriately assume it and should mention it clearly

- | | (Marks) |
|--|----------------|
| Q.1. a) Explain K-means clustering with example. | (6) |
| b) Explain following terms: Support, Confidence and Lift. Illustrate these terms with the help of a suitable example . | (6) |
| Q.2. a) What is document term frequency matrix? How is it useful in statistical analysis? | (6) |
| b) What is text mining? Write R language code for: i) to convert lower case to upper case letters ii) Removing Punctuation iii) Removing numbers. | (6) |
| Q.3. a) What is regression? Which tools are available in R for regression analysis? Explain them. | (6) |
| b) What is correlation? What are the three types ? Enlist the tools which are available in R for Correlation. | (6) |
| Q.4 a) What is bivariate data? Explain function used to group and organize bivariate data. | (6) |
| b) Enlist packages used to provide mapping information in R. | (6) |
| Q.5. a) What is supervised and unsupervised machine learning? Explain these terms with real life examples. Enlist algorithm used in supervised and unsupervised machine learning. | (6) |
| b) What is predictive modeling? What are the applications of predictive modeling? What are the different Modelling methods? | (6) |
| Q.6. Write a note on any 3 of the following | |
| a) K- nearest neighbor classification | (4) |
| b) Bayesian Hierarchical Clustering | (4) |
| c) Word Stems | (4) |
| d) Anomaly Detection | (4) |

