



Avinashilingam Institute for Home Science and Higher Education for Women
(Deemed to be University under Category 'A' by MHRD, Estd. u/s 3 of UGC Act 1956)
Re-accredited with 'A++' Grade by NAAC. Recognised by UGC Under Section 12B
Coimbatore - 641 043, Tamil Nadu, India

Continuous Internal Assessment II – October 2025

21BCSC21 Fundamentals of Data Science

Semester V

Class : III UG (2022 Batch Repeated)
Major : Computer Science

Time : 2 Hrs

Max.Marks : 60

Course Outcomes:

- CO1. Students will be able to apply the basic Data Science knowledge on the day to day Problems they encounter.
CO 2. Students will realize that there are various phases that contribute to the completion of a Data Science Project and can select among the various modeling techniques.
CO3. Students will be able to apply Regression techniques for modeling a data science project.
CO4. Students will be able to apply the Clustering and Association rule mining for modeling a Data science project.
CO5. Students can reproduce the knowledge gained and come out with a sample case study which they come across in their daily life and implement, document and present the same using the R Tool.

Part A

6 x 1 = 6

Choose the Correct Answer

1. The function used for linear regression in R is _____. CO3 K1
a) lm (formula, data) b) lr(formula, data)
c) lrm (formula, data) d) rg.linear(formula,data)
2. Which of the following is finally produced by Hierarchical clustering? CO4 K1
a) final estimate of cluster centroids
b) tree showing how close things are to each other
c) assignment of each point to clusters
d) all of the mentioned
3. The unit of 'togetherness' when mining association rule is _____. CO4K2
a) itemset b) transaction c) basket d) group
4. What technique can be used to improve the efficiency of apriori algorithm? CO4K2
a) Sampling b) Cleaning c) Hash based technique d) Transaction increase
5. _____ is an R package that allows the inclusion of R code & results inside documents. CO5K2
a) Knitr b) Knit c) hclust d) ggplot
6. Data Scientist are interested in _____ approach of a data science project. CO5K2
a) Structural b) Modelling c) Technique d) Pattern-based

Part B

3 x 6 = 18

Answer ALL questions

Each answer should not exceed 400 words or two pages

7. a. Summarize the procedure on reading the model summary & characterizing the coefficient quality in Linear Regression. CO3K1
(or)
b. Highlight the features on Linear Regression. CO3K2
8. a. Explain the views on hierarchical clustering with hclust() in brief. CO4K2
(or)
b. Explain the steps for K-means algorithm Procedure with suitable example. CO4K3
9. a. Discuss about using knitr to produce milestone documentation. CO5K2
(or)
b. Explain about presenting the result to the project sponsor. CO5K2

Part C

3 x 12 = 36

Answer ALL questions

Each answer should not exceed 800 words or four pages

10. a. Explain the significance and procedure on mining association rules with examples. CO4K1
(or)
b. Elucidate on grouping of data with similar characteristics using cluster analysis. CO4K2
11. a. Explain in detail about deploying models. CO4K2
(or)