DEPARTMENT OF STATISTICS				*Certificate courses (PG students)*				
Sem	**Course Type**	**Course Code**	**Course Title**	**Credits**	**Total Contact Hours**	**CIA**	**Ext**	**Total**
	Certificate Course		Cluster analysis with R	2	30	50	50	100

Nature of Course				
Knowledge and skill	✓		Employability oriented	
Skill oriented	✓		Entrepreneurship oriented	

Course Objectives:

The main objectives of this course are to:

1. Comfortably perform basics operations in R
2. Understand cluster analysis concepts
3. Explore and execute the cluster analysis concepts for real time data using R

Unit	Description	Hours
1	Basic of R: Introduction to R – Installing R software – loading packages – Getting Help – importing your data into R – data Preparation - Package R required.	**6**
2	Clustering Distance measure: Method for measuring distances – Choosing distance measure – data standardization – Distance matrix computation – visualizing distance matrices.	**6**
3	Partitioning Clustering: K – Means Clustering – basic – algorithm – Computing K – means clustering in R – Advantages and disadvantages – K-Medoids – CLARA.	**6**
4	Hierarchical Clustering: Agglomerative Clustering - Comparing dendrograms – visualizing dendrograms – Headmaps.	**6**
5	Cluster validation: Assessing cluster tendency – determining the optimal number of cluster – Cluster validation statistics – choosing the best clustering algorithms – computing P-values for Hierarchical clustering.	**6**
	Total number of hours	**30**

Books for study:

1. AlboukadelKassambara (2017), MultivariateAnalysis – I:Practical Guide To Cluster Analysis in R Unsupervised Machine learning, STHDA.

Book for references:

1. Giordani, paolo, Ferraro mariabrigida, Martella, Francesca (2019),An Introduction to Clustering with R, Springers**.**
2. Brain S. Evertti, Sabine Landu, MorvenLeese and Daniel Stahl (2011), John Wiley and Sons Ltd, 5 th editions, United Kingdom.