

GANPAT UNIVERSITY									
FACULTY OF ENGINEERING S TECHNOLOGY									
Programme		Bachelor of Technology			Branch /Spec.	Computer Science & Engineering (CSE) and CSE(BDA/CS/AI & ML)			
Semester		V			Version	1.0.0.1			
Effective from Academic Year				2026-27		Effective for the batch Admitted in			June 2024
Subject code		2CSE50E27		Subject Name		DATA SCIENCE S MODELING USING R			
Teaching scheme					Examination scheme (Marks)				
(Per week)	Lecture(DT)		Practical (Lab.)		Total		CE	SEE	Total
	L	TU	P	TW					
Credit	3	0	1	0	4	Theory	40	60	100
Hours	3	0	2	0	5	Practical	30	20	50
Pre-requisites:									
Probability S statistics, Probability distribution									
Learning Outcome:									
After completion of the course, student will be able to:									
<ul style="list-style-type: none"> • Learn the fundamentals of data analytics and the data science stream • Apply statistical methods, regression techniques and related algorithms to both large and small data sets in R Programming for prediction. • Demonstrate knowledge of statistical data analysis techniques utilized in decision making. • Implement hypothesis testing, various algorithms using various software platforms. 									
Theory syllabus									
Unit	Content								Hrs
1	Descriptive and Inferential Statistics Descriptive Statistics, Inferential Statistics through hypothesis tests								8
2	ANOVA S correlation coefficient ANOVA (Analysis of Variance), Coefficient of correlation,								7
3	Optimization Introduction to optimization,Constrained optimization, Unconstrained optimization,Linear optimization, Gradient-based methods								8
4	Regression Differentiating algorithmic and model based frameworks, Regression: Ordinary Least Squares,Ridge Regression, Lasso Regression, Logistic Regression K Nearest Neighbours,								8

5	Fundamentals of R: R Data Structures, Common Vector operations, Matrices, Arrays, Lists and Data Frames.	7
6	Reproducible Research Using R Reproducible Research using R and Rstudio (knitr, rmarkdown, bookdown, interactive document, shiny presentation, shiny web application)	7
Self-Study Topics		
<ul style="list-style-type: none"> Two-way ANOVA. 		

Practical Content
<ul style="list-style-type: none"> Practical's are to be designed in such a way that students must help students understand various types of regression models. The labs must emphasize on practical aspects of clustering, rule mining, hurdles dusting big data analytics. The labs must focus on active learning of practical concepts so as to achieve the course outcome with effectiveness.

Suggested Practical List
<p>practicals will be based on following criteria</p> <ul style="list-style-type: none"> Initial practical will introduce programming tools and its feature/power of implementation of problems. Practical related to inferential statistics and various hypothetical tests are implemented. Practical related to advanced topics like regression, clustering and classification are implemented with various aspects. Covered topics will create a concrete platform to advance techniques of machine learning, deep learning, reinforcement learning. <p>Suggested Softwares: R Programming, Excel. SPSS</p>

Mooc Course
<p>Course Name: Data Science for Engineers Link: https://onlinecourses.nptel.ac.in/noc20_cs72/course</p>

Text Books

1	Hastie, Trevor, et al. The elements of statistical learning. springer,
---	--

Reference Books

1	Montgomery, Douglas C., and George C. Runger. Applied statistics and probability for engineers. John Wiley S Sons, 2010
---	---

2	Bekkerman et al. Scaling up Machine Learning
---	--

3	Research Methodology Methods S Techniques, C. R. Kothari, Second Edition,2009
---	---

4	Vincent Granville, Developing Analytic Talent: Becoming a Data Scientist, wiley, 2014.
---	--

CourseOutcomes:

COs	Description
CO1	Learn the fundamentals of data analytics and the data science stream
CO2	Apply statistical methods, regression techniques and related algorithms to both large and small data sets in R Programming for prediction.
CO3	Demonstrate knowledge of statistical data analysis techniques utilized in decision making.
CO4	Implement hypothesis testing, various algorithms using various software platforms.

Mapping of CO and PO:

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	2	2	1	1	2	2	2	2	1	2	3	3
CO2	3	3	2	2	3	3	1	1	1	3	2	2
CO3	3	2	2	3	3	1	2	3	2	1	1	3
CO4	3	3	2	2	3	3	1	2	1	2	2	2