

Paper Title	MACHINE LEARNING USING R PROGRAMMING
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

To make the students learn the statistics & mathematical concepts, Hypothesis & Dimension Reduction Technique, R Programming Concepts and Machine Learning.

COURSE OUTCOMES:

CO1: Understand the fundamental concepts of Statistics & Mathematics

CO2: Understand Hypothesis & Dimension Reduction Techniques

CO3: Hands on Experience in R Programming

CO4: Understand Machine Learning Concepts using R

CO5: To have basic knowledge of various predictive models.

UNIT 1: STATISTICS & MATHEMATICAL ESSENTIALS

9 Hrs.

Measure of Central Tendency - Mean, Median, Mode - Dispersion Technique - Range Inter Quartile Range - Variance, Standard Deviation - Mean Square Error & Root Mean Square - Probability Distribution.

UNIT 2: HYPOTHESIS AND DIMENSION REDUCTION TECHNIQUE

9 Hrs.

Types of Hypothesis - Sample testing - T-test - Z-test - Chi-square test - Anova test -. One Way Anova. Two Way Anova - Principle component analysis - Collinearity and multicollinearity

UNIT 3: R PROGRAMMING CONCEPTS

9 Hrs.

The Data types in R & its uses -Build in functions in R- Data Manipulation - Data import Techniques – Exploratory Data Analysis – Data Visualization.

UNIT 4: MACHINE LEARNING

9 Hrs.

ML Fundamental & common use cases - Approach to Machine Learning - Understanding Supervised learning technique - Unsupervised learning technique

UNIT 5: PREDICTIVE MODELLING IN R

9 Hrs.

Introduction to predictive modeling - Regression Problem - Classification Problem - Linear Regression - Logistic Regression – Clustering - Distance measure types- K means clustering – Decision Tree Classifier – Random Forest Classifier – Support Vector Machine.

TEXTBOOK:

1. Introducing Data Science, Davy Cielen, Arno D. B. Meysman and Mohamed Ali, Manning Publications, 2016.
2. Think Like a Data Scientist, Brian Godsey, Manning Publications, 2017.

SUGGESTED BOOK:

1. Doing Data Science, Straight Talk from the Frontline, Cathy O'Neil, Rachel Schutt, O' Reilly, 1st edition, 2013.
2. Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, Cambridge University Press, 2nd edition, 2014
3. An Introduction to Statistical Learning: with Applications in R, Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer, 1st edition, 2013