MANAV RACHNA INTERNATIONAL INSTITUTE OF RESEARCH AND STUDIES
(Deemed to be University under section 3 of the UGC Act 1956)
NAAC 'A' Grade University

BCS-DS-727: Data Science

Periods/week	Credits	Max. Marks	: 200
L :3 T: 0	3.0	Continuous Evaluation	: 100
Duration of Exam: 3 Hrs		End Sem Examination	: 100

Pre-Requisite: Basic Knowledge of computers
Course Type: Program Electives

Course Outcomes: The students will be able to-
BCS-DS-727.1 Describe a flow process for data science problems.
BCS-DS-727.2 Classify data science problems into standard typology.
BCS-DS-727.3 Correlate results to the solution approach followed.
BCS-DS-727.4 Construct use cases to validate approach and identify modifications required.
BCS-DS-727.5 Develop R codes for data science solutions.
BCS-DS-727.4 Assess the solution approach.

PART –A

Unit-1. INTRODUCTION TO DATA SCIENCE:
1.1 Data science for engineers Course philosophy and expectation
1.2 Introduction to R
1.3 Variables and data types in R, Data frames
1.4 Recasting and joining of data frames
1.5 Arithmetic, Logical and Matrix operations in R
1.6 Control structures, Data visualization in R Basic graphics The study of language, applications of NLP

Unit-2. LINEAR ALGEBRA AND STATISTICAL MODELLING FOR DATA SCIENCE
2.1 Linear Algebra
2.2 Solving Linear Equation
2.3 Linear Algebra - Distance, Hyperplanes and Halfspaces, Eigen values, Eigen vectors
2.4 Statistical Modelling
2.5 Random Variables and Probability Mass/Density Functions
2.6 Sample Statistics
2.7 Hypotheses Testing

Unit-3. Optimization for Data Science
3.1 Basics of Optimization
3.2 Unconstrained Multivariate Optimization
3.3. Unconstrained Multivariate Optimization
3.4 Gradient (Steepest) Descent (OR) Learning Rule
3.5 Multivariate Optimization With Equality Constraints
3.6 Multivariate Optimization With Inequality Constraints

PART –B

Unit-4. PREDICTIVE MODELLING AND REGRESSION FOR DATA SCIENCE
4.1 Module : Predictive Modelling
4.2 Linear Regression
4.3 Model Assessment
4.4 Diagnostics to Improve Linear Model Fit
4.5 Simple Linear Regression Model Building
4.6 Simple Linear Regression Model Assessment

4.7 Simple Linear Regression Model Assessment (Continued)
4.8 Muliple Linear Regression

Unit-5. VALIDATION AND LOGIGITIC REGRESSION
5.1 Cross Validation
5.2 Multiple Linear Regression Modeling Building and Selection
5.3 Classification
5.4 Logisitic Regression
5.5 Performance Measures
5.6 Logisitic Regression Implementation in R

Unit-6. KNN and Recent trends in various data collection and analysis techniques
6.1 K - Nearest Neighbors (kNN)
6.2 K - Nearest Neighbors implementation in R
6.3 K - means Clustering
6.4 K - means implementation in R
6.5 Recent trends in various data collection
6.6 Analysis techniques
7.7 Various visualization techniques
7.8 Application development methods of used in data science.

Text/Reference Books:
1. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk From The Frontline. O'Reilly.
2. Jure Leskovek, AnandRajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press.

Instructions for paper setting: Seven questions are to be set in total. First question will be conceptual covering entire syllabus and will be compulsory to attempt. Three questions will be set from each Part A and Part B (one from each unit) Student needs to attempt two questions out of three from each part. Each question will be of 20 marks.

Distribution of Continuous Evaluation:

Sessional- I	30%
Sessional- II	30%
Assignment/Tutorial	20%
Class Work/ Performance	10%
Attendance	10%

Evaluation Tools:
Assignment/Tutorials
Sessional Tests
Surprise questions during lectures/Class Performance
End Sem Examination

COURSE ARTICULATION MATRIX :

CO Statement (BCS-DS-727)	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
BCS-DS-727.1	3	2	1	1	1	1	1	2	1	1	1	3	2	1	1
BCS-DS-727.2	1	3	1	1	3	1	1	1	1	1	1	2	1	1	1
BCS-DS-727.3	2	1	2	1	1	3	3	1	1	1	1	3	1	1	1
BCS-DS-727.4	1	2	2	3	1	2	2	1	1	2	1	2	2	1	2
BCS-DS-727.5	1	1	2	3	3	1	1	2	2	2	1	2	1	1	1
BCS-DS-727.6	1	1	3	1	2	1	1	2	3	1	2	2	2	3	2