Ph.D Semester 1/2
(CO / IT Engineering)

Subject Name: Data Science for Engineers **Subject Code:** PTCO13104
Type of course: PhD Course Work
Prerequisites (if any): - Statistics

Rationale: This course introduces the mathematical foundations required for data science. As part of this course, students will learn the framework for solving problems requiring data analytics. This course also introduces 'R' programming language. Students will be able to solve predictive problems on real time data after learning this course.

Teaching and Examination Scheme:

Teaching Scheme				Theory Marks			Practical Marks		Total
L	T	P	C	TEE	CA1	CA2	TEP	CA3	
4	0	0	4	60	25	15	0	0	100

CA1: Continuous Assessment (assignments/projects/open book tests/closed book tests) CA2: Sincerity in attending classes/class tests/ timely submissions of assignments/self-learning attitude/solving advanced problems TEE: Term End Examination TEP: Term End Practical Exam (Performance and viva on practical skills learned in course) CA3: Regular submission of Lab work/Quality of work submitted/Active participation in lab sessions/viva on practical skills learned in course

Contents

Sr No	Description	No. of Hours
1.	Course philosophy and expectation, Introduction to R, Variables and data types in R, Data frames, Recasting and joining of data frames, Arithmetic, Logical and Matrix operations in R, Advanced programming in R : Functions, Control structures, Data visualization in R Basic graphics	8
2.	Linear Algebra for Data science, Solving Linear Equations, Linear Algebra - Distance, Hyperplanes and Halfspaces, Eigenvalues, Eigenvectors	6
3.	Statistical Modelling, Random Variables and Probability Mass/Density Functions, Sample Statistics, Hypotheses Testing	6
4.	Optimization for Data Science, Unconstrained Multivariate Optimization, Gradient Descent, Learning Rule	6

5.	Multivariate Optimization With Equality Constraints, Multivariate Optimization With Inequality Constraints, Introduction to Data Science, Solving Data Analysis Problems - A Guided Thought Process	6
6.	Module: Predictive Modelling, Linear Regression, Model Assessment, Diagnostics to Improve Linear Model Fit, Simple Linear Regression Model Building, Simple Linear Regression Model Assessment, Multiple Linear Regression	6
7.	Cross Validation, Multiple Linear Regression Modelling Building and Selection, Classification, Logistic Regression, Performance Measures, Logistic Regression Implementation in R	6
8.	K - Nearest Neighbors (kNN), K - Nearest Neighbors implementation in R, K - means Clustering, K - means implementation in R	6
9.	Applications of Data Science, Challenges reported in contemporary literature, Case Studies	10

Course Outcomes

Sr. No.	CO Statement	Marks % Weightage
1.	Describe a flow process for data science problems	20%
2.	Classify data science problems into standard typology	20%
3.	Develop R codes for data science solutions	20%
4.	Correlate results to the solution approach	15%
5.	Assess the solution approach	15%
6.	Construct use cases to validate approach and identify modifications required	10%

Reference Books:

Sr no	Title of book /article	Author(s)	Publisher	Year of publication	Publication Edition
1	Introduction To Linear Algebra	Gilbert Strang	Wellesley Publishers	2016	Fifth Edition
2	Applied Statistics And Probability For Engineers	Douglas Montgomery	Wiley	2016	Sixth Edition

3	Practical Statistics for Data Scientists	Peter Bruce and Andrew Bruce	Publisher(s): O'Reilly Media, Inc. ISBN: 9781491952962	2017	2nd Edition
4	Think Stats: Probability and Statistics for Programmers	Allen B. Downey	Green Tea Press Needham, Massachusetts	2014	2nd Edition
5	Data Science for Dummies	John Muller	Wiley	2017	2nd Edition
6	Statistics for Data Science	James D. Miller	PACKT	2017	-

List of Online Learning Resources:

3. NPTEL course on: Data Science for Engineers (https://onlinecourses.nptel.ac.in/noc23_cs17/preview)