Mathematical Foundation for Data Science

Code: CS32111CS	Course Type: Core	Semester: I
Evolution Scheme (TA-MSE-ESE): 20-30-100	Total Marks: 150	L-T-P (Credits): 3-0-0 (3)

Pre-Requisites: Calculus, Linear Algebra, Probability & Statistics

Course Objective: To provide a strong mathematical basis for data science by integrating linear algebra, probability, statistics, and optimization techniques required for analyzing high-dimensional data and building robust machine learning models.

Course Outcomes: At the end of the course, the student will be able to

CO-1	Apply linear algebra techniques for data representation, dimensionality reduction, and feature extraction
CO-2	Utilize probability theory and statistical methods for data analysis and inference
CO-3	Formulate and solve optimization problems underlying machine learning algorithms
CO-4	Analyze uncertainty and variability in data using stochastic processes and statistical modeling

Course Articulation Matrix:

PO	PEO-1	PEO-2	PEO-3	PEO-4	PEO-5
CO-1	3	2	1	-	-
CO-2	3	2	1	-	1
CO-3	3	2	-	-	3
CO-4	3	2	1	1	-

1 - Slightly; 2 - Moderately; 3 – Substantially

Syllabus:

Unit-1: Linear Algebra for Data Science: Vector spaces, norms, inner products, orthogonality, matrix operations, eigen decomposition, singular value decomposition (SVD), principal component analysis (PCA), dimensionality reduction techniques.

Unit-2: Probability Theory: Probability axioms, conditional probability, Bayes theorem, random variables, joint and marginal distributions, expectation, variance, covariance, commonly used distributions.

Unit-3: Statistical Methods: Estimation theory, maximum likelihood estimation, hypothesis testing, confidence intervals, regression analysis, multivariate statistics, sampling techniques, statistical inference in data science.

Unit-4: Optimization and Numerical Methods: Gradient descent and its variants, convex optimization, constrained optimization using Lagrange multipliers, numerical stability, optimization in machine learning models.

Learning Resources:

Textbooks:

1. Deisenroth et al., Mathematics for Machine Learning
2. Hastie et al., Elements of Statistical Learning
3. Boyd & Vandenberghe, Convex Optimization