Course Code	Course Name	Teaching Scheme (Contact Hours)			Credits Assigned			
		Theory	Prac	Tut	Theory	Prac	Tut	Total
MEAIC103	Mathematical Foundations of Data Science	04	-	-	04	-	-	04

Examination Scheme							
Theory					TW	Oral/ Prac	Total
Internal			End Sem. Exam	Exam Duration (in Hrs)			
Test 1	Test 2	Avg					
20	20	20	80	03	-	-	100

Course Objectives:

This course will introduce students to the fundamental mathematical concepts required for applying data science.

Course Outcomes:

Upon completion of the course, the learners will be able to:

1. Understand the importance of linear algebra, statistics and probability from data science perspective.
2. Understand the elements of structured data and data distribution for binary as well as categorical data.
3. Apply the knowledge of sampling and distribution algorithms to evaluate the real distribution of sampling data.
4. Apply the knowledge of significance testing, use of null value hypothesis to outline the conditions for a particular test.
5. Evaluate and analyze the results of confusion matrix.
6. Apply optimization techniques for improvising performance.

Prerequisites: Fundamentals of Probability and Statistics.

Sr. No.	Module	Detailed Content	Hours
1	Basics of Data Science	Introduction; Importance of linear algebra, statistics and optimization from a data science perspective; Structured thinking for solving data science problems; Probability, Statistics and Random Processes: Probability theory and axioms; Random variables.	8
2	Linear Algebra	Matrices and their properties (determinants, traces, rank, nullity, etc.); Eigenvalues and eigenvectors; Matrix factorizations; Inner products; Distance measures.	8
3	Exploratory Data Analysis	Elements of structured data; Estimates of location; Estimates of variability; Expectations and moments; Exploring the data distribution; Exploring binary and categorical data; Covariance and correlation; Exploring two or more variables.	8
4	Data and Sampling Distributions	Random sampling and sample bias; Selection bias; Central limit theorem; Standard error; Bootstrap;	8

		Confidence intervals; Normal distribution; Long-tailed distributions; Student's t-distribution; Binomial distribution; Poisson distributions; Exponential distribution; Weibull distribution; Fitting a model.	
5	Statistics and Significance Testing	Hypothesis tests; A/B testing; Chi-square test; confidence intervals; p-values; ANOVA; t-test; Confidence (statistical) intervals; Degrees of freedom; White-noise process.	8
6	Evaluation and Optimization	Mathematics in algorithmic performance evaluation: Confusion matrix; Precision; Recall; Specificity; ROC Curve; AUC; Lift; Optimization: Global and local optima; Unconstrained and constrained optimization; Introduction to least squares optimization.	8

Text Books:

1. P. Bruce and A. Bruce, Practical Statistics for Data Scientists: 50 Essential Concepts, O'Reilly.
2. C. O'Neil and R. Schutt, Doing Data Science, O'Reilly.

Reference Books:

1. G. Strang, Introduction to Linear Algebra, 5th edition, Wellesley-Cambridge Press, USA.
2. W. Hines, D. Montgomery, D. Goldman, C. Borror, Probability and Statistics in Engineering, Wiley India Pvt. Ltd.
3. A. Agresti, C. Franklin, B. Klingenberg, Statistics: The Art and Science of Learning from Data, Global Edition, Pearson.

Internal Assessment:

Assessment consists of two tests out of which one should be compulsorily class test (on minimum 02 modules) and the other can be either a class test or assignment on real-world problems or course related project.

Theory Examination:

1. Question paper will comprise of total 6 questions.
2. All questions carry equal marks.
3. Questions will be mixed in nature (for example, suppose Q2 has part (a) from module 3, then Q2 part (b) will be from any module other than module 3).
4. Only 4 questions need to be solved.
5. In question paper, weightage of each module will be proportional to the number of respective lecture hours as mentioned in the syllabus.