|  |  |
| --- | --- |
| Program: Bachelor of Science Honours (Information Technology) | Semester: II |
| Course: Statistical Methods for Data Science | Code:  |
| Teaching Scheme | Evaluation Scheme |
| Lecture | Practical | Tutorial | Credits | Theory | Practical |
| Internal | External | Internal | External |
| 30 | Nil | Nil | 02 | 20 Marks  |  30 Marks | Nil | Nil |
|   |
| Internal Component  |
| Class Test Duration Mins | Assignment& projects |  Class Participation |
|  10 Marks |  10 Marks | Nil |
|   |
| Learning Objectives1. To understand the concepts and techniques used in numerical methods for solving 1st and 2nd order differential equations.
2. Understand the concept of correlation and its importance in analyzing the relationship between variables
3. Understand the concepts of discrete and continuous random variables and their probability distributions
 |
| Learning Outcomes* Taylor series expansion to approximate solutions of differential equations.
* Implementation Euler's Method, Modified Euler's Method, and Runge-Kutta Method for solving 1st and 2nd order differential equations
* Apply least-squares regression techniques to fit linear, polynomial, and multiple linear regression models.
* Perform correlation analysis and interpret the results to make meaningful conclusions about the relationships between variables
* Data science concepts, including an understanding of big data, statistical inference, exploratory data analysis, and basic machine learning algorithms like linear regression, k-nearest neighbours (k-NN), and k-means.
 |
| Pedagogy* PPTs, Case studies, Group discussions, Classroom Activity, Videos, Research papers, News articles etc.
 |

**Module 1** (10)

Numerical solution of 1st and 2nd order differential equations: Taylor series, Euler’s Method, Modified Euler’s Method, Runge-Kutta Method for 1st and 2 and Order Differential Equations

Correlation Analysis, Least-Squares Regression:

Linear Regression, Polynomial Regression, Multiple Linear Regression, General Linear Least Squares, Nonlinear Regression

**Module 2** (10)

Random variables: Discrete and Continuous random variables, Probability density function, Probability distribution of random variables, Expected value, Variance.

Distributions: Discrete distributions: Uniform, Binomial, Poisson, Bernoulli, Continuous distributions: uniform distributions, exponential, (derivation of mean and variance only and state other properties and discuss their applications) Normal distribution state all the properties and its applications.

**Module 3** (10)

Introduction: What is Data Science? - Big Data and Data Science hype – and getting past the hype - Why now? – Datafication - Current landscape of perspectives - Skill sets needed.

Statistical Inference - Populations and samples - Statistical modeling, probability distributions, fitting a model - Intro to R

Exploratory Data Analysis and the Data Science Process - Basic tools (plots, graphs and summary statistics) of EDA - Philosophy of EDA - The Data

Science Process. Three Basic Machine Learning Algorithms - Linear Regression - k-Nearest Neighbors (k-NN) - k-means