DSE-2B(2)

Course code: CS-622 T **Course Title: B2. Advance Data Science**
Total Credit: 2 Marks: 50 (UA: 40 + IA: 10)
Periods: 3 per week (50 Minutes each)

Prerequisites : Basic data Science

Learning Objectives

.Data Science is a fast-growing interdisciplinary field, focusing on the analysis of data to extract knowledge and insight.
.To introduce students to the collection, Preparation, analysis, modelling and visualization of data, covering both conceptual and practical issues.
.Examples and case studies from diverse fields will be presented, and hands-on use of statistical and data manipulation software will be included

Learning Outcomes
After learning the course the students should be able to:
.Recognize various disciplines that contribute to a successful data science effort.
.Identify the data types, relation between data and visualization technique for data.
.Explain probability, distribution, sampling, Estimation
.Solve regression and classification problem.
.Develop and appreciate various techniques for data modelling and mining.

Course Outline :
Unit – 1 Introduction to Probability, Introduction to Probability Theory, Probability Theory – Terminology, Fundamental Concepts in Probability – Axioms of Probability, Application of Simple Probability Rules – Association Rule Learning, Bayes' Theorem, Random Variables Probability Density Function (PDF) and Cumulative Distribution Function (CDF) of a Continuous Random Variable, Binomial Distribution, Poisson Distribution, Geometric Distribution, Parameters of Continuous Distributions, Uniform Distribution, Exponential Distribution, Chi-Square Distribution, Student's t-Distribution, F-Distribution.
Unit – 2 Sampling and Estimation, Introduction to Sampling, Population Parameters and Sample Statistic, Sampling, Probabilistic Sampling, Non-Probability Sampling, Sampling Distribution, Central Limit Theorem (CLT), Sample Size Estimation for Mean of the Population, Estimation of Population Parameters, Method of Moments, Estimation of Parameters Using Method of Moments, Estimation of Parameters Using Maximum Likelihood Estimation,

Unit – 3 simple Linear Regression, Introduction to Simple Linear Regression, History of Regression–Francis Galton's Regression Model, Simple Linear Regression Model Building, Estimation of Parameters Using Ordinary Least Squares, Interpretation of Simple Linear Regression Coefficients, Validation of the Simple Linear Regression Model,Outlier Analysis, Confidence Interval for Regression Coefficients b0 and b, Confidence Interval for the Expected Value of Y for a Given X, Prediction Interval for the Value of Y for a Given X.

Unit – 4 Logistic Regression, Introduction – Classification Problems, Introduction to Binary Logistic Regression, Estimation of Parameters in Logistic Regression, Interpretation of Logistic Regression Parameters, Logistic Regression Model, Diagnostics, Classification Table, Sensitivity, and Specificity, Optimal Cut-Off Probability, Variable Selection in Logistic Regression, Application of Logistic Regression in Credit Rating, Gain Chart and Lift Chart

Unit - 5 Test and Tutorial

Reference Books:-

1. Dinesh Kumar, Business Analytics, Wiley IndiaBusinenalytics: The Science
2. V.K. Jain, Data Science & Analytics, Khanna Book Publishing, New Delhi of Dat
3. Data Science For Dummies by Lillian Pierson , Jake Porway
4) Doing Data Science by Rachel Schutt, Cathy O'Neil, O'Reilly publication
5) Data Science with Jupyter Author: Prateek Gupta, BPB publication