Course code: Course Title	Course Structure			Pre-Requisite
SE329: Methods for Data Analysis	L	T	P	NIL
	3	0	2	

Course Objective: To make one understand the methods for data preparation and analysis.

S. NO	Course Outcomes (CO)
CO1	Understand the principles and importance of data analysis, including effective data collection strategies and mining software repositories.
CO2	Identify different types of variables, and classify data using appropriate measurement scales.
CO3	Apply descriptive statistics techniques to summarize data and inferential statistics methods to draw meaningful conclusions.
CO4	Implement data preparation techniques such as feature selection, and feature extraction in order to have quality data for model development.
CO5	Apply various data analysis techniques, including statistical and machine learning methods, to analyze data effectively and solve real-world problems.

S. NO	Contents	Contact Hours
UNIT 1	**Introduction:** Data Collection Strategies, Data Collection from Repositories, Mining Data from Software Repositories: Configuration Management Systems, Importance of Mining Software Repositories. Common Types of Software Repositories, Version Control Systems, Bug Tracking Systems, Open Source Repositories.	6
UNIT 2	**Types of Variables:** Independent and Dependent Variables, Categorical vs Numerical, Nominal Variables, Ordinal Variables, Interval Variables, Ratio Variables; Identifying the dependent and independent variables, Confidence levels.	8
UNIT 3	**Data Preparation-I:** Descriptive Statistics: Summarizing and describing a collection of data, Univariate and bivariate analysis, Mean, mode and standard deviation, Percentages and Ratios, Histograms, Identifying randomness and uncertainty in data inferential Statistics: Drawing inference from data, Modeling assumptions, Identifying Patterns, Regression analysis, T-test, Analysis of Variance, Correlations, Chi-square Measures of central tendency, measures of dispersion, data distribution, histogram analysis, normalization, outlier analysis, correlation analysis.	6
UNIT 4	**Data Preparation-II**: Attribute Reduction Methods: Univariate Analysis, Correlation-based Feature Selection, Attribute Extraction: Principal Component Analysis.	6
UNIT 5	**Data Analysis:** Data Analysis Techniques: Introduction to Statistical and Machine Learning techniques, Tools for analyzing Data.	8
UNIT 6	**Applications:** Case studies for data preparation and analysis.	8
	TOTAL	42

REFERENCES

S.No.	Name of Books/Authors/Publishers	Year of Publication / Reprint
1	Max Kuhn, Kjell Johnson, "Applied Predictive Modelling", Springer, 2nd Edition.	**2018**
2	Ruchika Malhotra, "Empirical Research in Software Engineering: Concepts, Analysis & Applications", CRC press, 1st Edition.	**2016**
3	Kattamuri S. Sarma, "Predictive Modeling with SAS Enterprise Miner: Practical Solutions for Business Applications", SAS Institute, 3rd Edition.	**2017**
4	Jeffrey Strickland, "Predictive Modeling and Analytics", Lulu.com.	**2014**