Title of Course: Data Analytics Systems and Algorithms **Course Code: 18B11CI917**
L-T-P Scheme: 3-0-0 **Credits: 3**

Pre-requisite:

- Students must have the minimal concept of Data Base Management Systems
- They must also have the concept of different types of algorithms used for searching data
- Students must have already studied the course "Business Analysis Techniques"

COURSE OVERVIEW:

This course will introduce students to this rapidly growing field of data analytics and equip them with some of its basic principles and tools as well as its general mindset. Students will learn concepts, techniques and tools they need to deal with various facets of data analytics practice, including data collection and integration, exploratory data analysis, predictive modelling, descriptive modelling, data product creation, evaluation, and effective communication.

Objective: The primary aim of this course is to further expand your understanding of data analytics and algorithms. To understand Data Analytics Life Cycle and Business Challenges. To understand Analytical Techniques and Statically Models

Learning Outcomes

Course Outcome	Description
CO1	**Demonstrate** proficiency with statistical analysis of data.
CO2	**Understand** the ability to build and assess data-based models.
CO3	**Demonstrate** skill in data management
CO4	**Illustrate** statistical analyses with professional statistical software.
CO5	**Implement** clustering algorithms like hierarchical Agglomerative clustering and k-means algorithm.
CO6	**Apply** data analytics concepts and methods to solve problems in real-world contexts and will communicate these solutions effectively

Course Outline:

Unit I: Introduction and Data Pre-processing: Data Science Introduction, Big Data and Data Science, Current landscape of perspectives

Unit II: Data Analysis and Correlations: Basic Concepts and Methods Populations and samples, Statistical modelling, probability distributions, Regression, fitting a model, Dimensionality Reduction: PCA & DWT, Correlation and regression analysis. Chi-square t and F distributions (definitions only) Confidence interval Single mean and difference known and unknown variances.

Unit III: Introduction to machine learning and Cluster Analysis: Basic Concept and Methods Supervised and unsupervised learning, Training and testing data, over fitting and under fitting. Distance measures: - Manhattan, Chebbychev, Mahalanobis Distance Cluster Analysis, Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid-Based Methods, Evaluation of Clustering, Clustering High-Dimensional Data Clustering Graph and Network Data

Unit IV: Classification Algorithms: Basic Concepts, Decision Tree Induction, Bayes Classification Methods, Rule-Based Classification, Model Evaluation and Selection, Techniques to Improve Classification Accuracy, Support Vector Machines, Lazy Learners (or Learning from Your Neighbors)

Unit V: Introduction to Web Search and Social Media Analytics: Data Wrangling: APIs and other tools for scrapping the Web Mining Complex Data Types, Other Methodologies of Data, Mining, Data Mining Applications, Data Mining and Society, Data Mining Trends Social Media Analytics is the science of analyzing data to convert information to useful knowledge. This knowledge could help us understand our world better and, in many contexts, enable us to make better decisions.

Evaluation Scheme:

Evaluations	Marks	Remarks
T-1	15 Marks (1-Hours)	1^{st} - 4^{th} Week
T-2	25 Marks (1:30 Hours)	5^{th} - 10^{th} Week
T-3	35 Marks (2-Hours)	11^{th} - 16^{th} Week
Assignments	10 Marks	
Tutorials / Subject Seminar	5 Marks	
Quiz	5 Marks	
Attendance	5 Marks	
Total	**100 Marks**	

Text Book

1. Data Mining, Concepts and Techniques: Jiawei Han and Micheline Kamber, Elsevier 2nd edition.
2. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk from the Frontline, O'Reilly. 2014.
3. Avrim Blum, John Hopcroft and Ravindran Kannan. Foundations of Data Science.

Refernce Books:

1. Data Mining: Introductory and Advanced Topics: Margaret H. Dunham, Prentice Hall.
2. Data Warehousing, Data Mining and OLAP: Alex, Berson, Stephen J. Smith, Tata McGraw- Hill, 2004.
3. Mining the Web : Discovering knowledge from hypertext data: Soumen Chakrabarty, Morgan Kaufmann
4. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press. 2014.
5. Trevor Hastie, Robert Tibshirani and Jerome Friedman. Elements of Statistical Learning, Second Edition. ISBN 0387952845. 2009.