UNIVERSITY OF MADRAS

B.Sc. DEGREE PROGRAMME IN COMPUTER SCIENCE

SYLLABUS WITH EFFECT FROM 2023-2024

Year: III **Semester: VI**

Big Data Analytics Common for B.C.A. , B.Sc.-SA , B.Sc.-CSc , B.Sc.-CSc-wAI , B.Sc.-CSc-wDS	**320E6E**
Credits 3	**Lecture Hours:5 per week**

Learning Objectives: (for teachers: what they have to do in the class/lab/field)

- To know the fundamental concepts of big data and analytics.
- To explore tools and practices for working with big data.

Course Outcomes: (for students: To know what they are going to learn)

CO1: Work with big data tools and its analysis techniques.
CO2: Analyse data by utilizing clustering and classification algorithms.
CO3: Learn and apply different mining algorithms and recommendation systems for large volumes of data.
CO4: Perform analytics on data streams.
CO5: Learn NoSQL databases and management.

Units	**Contents**
I	INTRODUCTION TO BIG DATA : Evolution of Big data — Best Practices for Big data Analytics — Big data characteristics — Validating — The Promotion of the Value of Big Data — Big Data Use Cases- Characteristics of Big Data Applications — Perception and Quantification of Value -Understanding Big Data Storage — A General Overview of High-Performance Architecture — HDFS — MapReduce and YARN — Map Reduce Programming Model
II	CLUSTERING AND CLASSIFICATION: Advanced Analytical Theory and Methods: Overview of Clustering — K-means — Use Cases — Overview of the Method — Determining the Number of Clusters — Diagnostics — Reasons to Choose and Cautions. - Classification: Decision Trees — Overview of a Decision Tree — The General Algorithm — Decision Tree Algorithms — Evaluating a Decision Tree — Decision Trees in R — Naïve Bayes — Bayes? Theorem — Naïve Bayes Classifier
III	ASSOCIATION AND RECOMMENDATION SYSTEM:Advanced Analytical Theory and Methods: Association Rules — Overview — Apriori Algorithm — Evaluation of Candidate Rules — Applications of Association Rules — Finding Association& finding similarity — Recommendation System: Collaborative Recommendation- Content Based Recommendation — Knowledge Based Recommendation- Hybrid Recommendation Approaches
IV	STREAM MEMORY: Introduction to Streams Concepts — Stream Data Model and Architecture — Stream Computing, Sampling Data in a Stream — Filtering Streams — Counting Distinct Elements in a Stream — Estimating moments — Counting oneness in a Window — Decaying Window — Real time Analytics Platform (RTAP) applications — Case Studies — Real Time Sentiment Analysis, Stock Market Predictions. Using Graph Analytics for Big Data: Graph Analytics

V	NOSQL DATA MANAGEMENT FOR BIG DATA AND VISUALIZATION: NoSQL Databases: Schema-less Models- Increasing Flexibility for Data Manipulation-Key Value Stores- Document Stores — Tabular Stores — Object Data Stores — Graph Databases Hive — Sharding —Hbase — Analyzing big data with twitter — Big data for E-Commerce Big data for blogs — Review of Basic Data Analytic Methods using R.

Learning Resources:

Recommended Texts

1. Anand Rajaraman and Jeffrey David Ullman, “Mining of Massive Datasets”, CambridgeUniversity Press, 2012.

Reference Books

1. David Loshin, “Big Data Analytics: From Strategic Planning to Enterprise Integration withTools, Techniques, NoSQL, and Graph”, Morgan Kaufmann/El sevier Publishers, 2013.

2. EMC Education Services, “Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data”, Wiley publishers, 2015.