DHANALAKSHMI SRINIVASAN ENGINEERING COLLEGE
(AUTONOMOUS)
(Approved by AICTE & Affiliated to Anna University, Chennai)
Accredited with 'A' Grade by NAAC, Accredited by TCS
Accredited by NBA with BME, ECE & EEE
PERAMBALUR - 621 212. Tamil Nadu.
website : www.dsengg.ac.in

COURSE PLAN (2025 – 2026 Odd Semesters)

Name of the Faculty				
Designation/Department	AP/IT			
Course Code/Name	U20IT701 / BIG DATA ANALYTICS			
Year/Section/Department	III / IT			
Credits Details	L: 3	T: 0	P: 0	C: 3
Total Contact Hours Required	45			

Syllabus:

UNIT I - INTRODUCTION TO BIG DATA	**No. of Periods:9**
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 HighPerformanceArchitecture-HDFS-MapReduceandYARN-MapReduceProgrammingModel	
UNIT II - CLUSTERING AND CLASSIFICATION	**No. of Periods: 9**
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.	
UNIT III - ASSOCIATION AND RECOMMENDATION SYSTEM	**No. of Periods: 9**
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.	
UNIT IV - STREAM MEMORY	**No. of Periods: 9**
Introduction to Streams Concepts – Stream Data Model and Architecture - Stream Computing, Sampling Data in a Stream – Filtering Streams – Counting Distinct Elements in a Stream – Estimatingmoments – Counting oneness in a Window – Decaying Window – Real time Analytics Platform(RTAP)applications - Case Studies - Real Time Sentiment Analysis, Stock Market Predictions. Using GraphAnalytics for Big Data: Graph Analytics	
UNIT V - NOSQL DATA MANAGEMENT FOR BIG DATA AND VISUALIZATION	**No. of Periods: 9**
NoSQL Databases : Schema-less Models‖: Increasing Flexibility for Data Manipulation-Key ValueStores- 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 BasicData Analytic Methods using R.	

Objective:

- To know the fundamental concepts of big data and analytics.
- To explore tools and practices for working with big data
- To learn about stream computing.
- To know about the research that requires the integration of large amounts of data.

Text Book:

T1. Anand Rajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, 2012.
T2. David Loshin, "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph", Morgan Kaufmann/El sevier Publishers, 2013

Reference Book:

R1. EMC Education Services, "Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data", Wiley publishers, 2015.
R2. Bart Baesens, "Analytics in a Big Data World: The Essential Guide to Data Science and its Applications", Wiley Publishers, 2015.
R3. Dietmar Jannach and Markus Zanker, "Recommender Systems: An Introduction", Cambridge University Press, 2010.
R4. Kim H. Pries and Robert Dunnigan, "Big Data Analytics: A Practical Guide for Managers " CRC Press, 2015.
R5. Jimmy Lin and Chris Dyer, "Data-Intensive Text Processing with MapReduce", Synthesis Lectures on Human Language Technologies, Vol. 3, No. 1, Pages 1-177, Morgan Claypool publishers, 2010.

Website:

W1. https://lecturenotes.in/notes/6099-notes-for-data-analytics-da-by-prasanta-bal
W2. https://nasrinword.wordpress.com/cp5293-big-data-analytics/

Online Mode of Study (if Any):

NPTEL details can be listed.

- https://www.coursera.org
- https://www.classcentral.com
- https://www.udemy.com

Course Plan:

Topic Number	Topic	Reference Detail	Page Number	Mode of teaching	Number of Periods Required	Cumulative Period
UNIT I - INTRODUCTION TO BIG DATA						**9**
1.	Evolution of Big data	T2	1	PPT	2	2
2.	Best Practices for Big data Analytics - Big data characteristics	T2	11-19	PPT	2	4
3.	Validating - The Promotion of the Value of Big Data	T2	20-24	BB	1	5
4.	Big Data Use Cases- Characteristics of Big Data Applications	T2	25-27	BB	1	6
5.	Perception and Quantification of Value	T2	40-42	BB	1	7
6.	Understanding Big Data Storage - A General Overview of High Performance Architecture	T2	44-52	BB	1	8
7	HDFS-Map Reduce and YARN-Map Reduce Programming Model	T2	55-63	BB	1	9
Outcome of Unit I: At the end of unit, Students should be able to • Identify big data use cases, characteristics and make use of HDFS and Map-reduce programming model for data analytics.						
UNIT II - CLUSTERING AND CLASSIFICATION						**9**
8	Advanced Analytical Theory and Methods: Overview of Clustering	T1	241	BB	1	10
9	K-means - Use Cases - Overview of the Method	T1	242-245	BB	1	11
10	Determining the Number of Clusters - Diagnostics	T1	247-249	BB	1	12
11	Reasons to Choose and Cautions	T1	251-256	BB	1	13
12	Classification: Decision Trees - Overview of a Decision Tree - The General Algorithm	T1	258-263	PPT	2	15
13	Decision Tree Algorithms - Evaluating a Decision Tree	T1	265-269	BB	2	17
14	Decision Trees in R - Naïve Bayes - Bayes‘ Theorem -Naïve Bayes Classifier.	T1	272	BB	1	18
Outcome of Unit II: At the end of unit, Students should be able to • Examine the data with clustering and classification techniques.						

UNIT III - ASSOCIATION AND RECOMMENDATION SYSTEM						**9**
15	Advanced Analytical Theory and Methods: Association Rules	T1	205	BB	2	20
16	Association Rules - Overview - Apriori Algorithm	T1	206-208	PPT	2	22
17	Evaluation of Candidate Rules - Applications of Association Rules	T1	209-210	BB	1	23
18	Finding Association& finding similarity	T1	211-212	BB	1	24
19	Recommendation System: Collaborative Recommendation	T1	307	BB	1	25
20	Content Based Recommendation	T1	308-310	BB	1	26
21	Knowledge Based Recommendation- Hybrid Recommendation Approaches	T1	312	PPT	1	27
Outcome of Unit III: At the end of unit, students should be able to • Discover the similarity of huge volume of data with association rule mining and examine recommender system.						
UNIT IV - STREAM MEMORY						**9**
22	Introduction to Streams Concepts – Stream Data Model and Architecture	T1	131	BB	1	28
23	Stream Computing, Sampling Data in a Stream – Filtering Streams	R3	205-208	BB	1	29
24	Counting Distinct Elements in a Stream – Estimating moments – Counting oneness in a Window	T1	136-138	BB	1	30
25	Decaying Window – RTAP applications	T1	139-142	PPT	2	32
26	Case Studies - Real Time Sentiment Analysis, Stock Market Predictions	T1	143-152	BB	2	34
27	Using Graph Analytics for Big Data: Graph Analytics	T1	153-167	BB	2	36
Outcome of Unit IV: At the end of unit, students should be able to • Explain analytics on data streams.						
UNIT V - NOSQL DATA MANAGEMENT FOR BIG DATA AND VISUALIZATION						**9**
28	NoSQL Databases : Schema-less Models	W1	---	BB	1	37
29	Increasing Flexibility for Data Manipulation	T2	84-85	BB	1	38
30	Key Value Stores- Document Stores - Tabular Stores - Object Data Stores	T2	86-87	BB	1	39
31	Object Data Stores - Graph Databases Hive - Sharding	R1	151-153	PPT	2	41
32	Hbase – Analyzing big data with twitter	T2	87-88	BB	2	43
33	Big data for E-Commerce Big data for blogs	R2	170-172	BB	1	44
34	Review of Basic Data Analytic Methods using R.	W2	---	PPT	1	45

Outcome of Unit V:

At the end of unit, Students should be able to

- Infer No SQL database and its management.
- Explain big data tools and its analysis techniques.

Course Outcome:

At the end of course:

Students should be able to do:

CO1: Identify big data use cases, characteristics and make use of HDFS and Map-reduce programming model for data analytics.(K2)

CO2: Examine the data with clustering and classification techniques.(K2)

CO 3: Discover the similarity of huge volume of data with association rule mining and examine recommender system.(K2)

CO 4: Explain analytics on data streams.(K2)

CO 5: Infer No SQL database and its management.(K2)

CO 6: Explain big data tools and its analysis techniques.(K2)

Course Outcome Vs Program Outcome Mapping:

Course outcomes	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	2	1	1	1	-	-	-	-	-	-	-	-	1	1
CO2	1	3	2	2	2	-	-	-	-	-	-	-	1	1
CO3	3	2	1	-	1	-	-	-	-	-	-	-	1	1
CO4	2	1	1	-	1	-	-	-	-	-	-	-	1	1
CO5	3	2	1	-	-	-	-	-	-	-	-	-	3	1
CO6	2	1	-	-	2	-	-	-	-	-	-	-	1	1
Avg:	**2.33**	**1.33**	**1.00**	**1.00**	**2.33**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1.33**	**1.00**

Content beyond Syllabus:

- SQL server and its applications
- Anatomy Of Map Reduce And Job Run
- Map Reduce Types And Formats
- Task Execution

Assignment:

Web portal	Assignment	Components	Topic Number with Topic / Unit Details	Relevance to CO
Web portal 1	--	Assessment – I (60)	Unit I and II	CO 1 & CO2
	1	Assignment – Handwritten (20)	4. Characteristics of Big Data Applications 7. HDFS-Map Reduce and YARN-Map Reduce Programming Model 9. K-means - Use Cases - Overview of the Method 14. Decision Trees in R - Naïve Bayes	CO1
	2	Assignment – Poster Presentation / PPT (20)	28. No SQL Databases 30. Key Value Stores- Document Stores - Tabular Stores - Object Data Stores 33. Big data for E-Commerce Big data for blogs 34. Data Analytic Methods using R.	CO2
Web portal 2	--	Assessment – II (60)	Unit III and IV	CO3 & CO4
	3	Seminar (20)	16. Association Rules - Overview - Apriori Algorithm 21.Knowledge Based Recommendation- Hybrid Recommendation Approaches 22. Streams Concepts – Stream Data Model and Architecture 25. RTAP applications	CO2
	4	Case Study Report (20)	26. Case Studies - Real Time Sentiment Analysis, Stock Market Predictions 32. Hbase – Analyzing big data with twitter	CO4
Web portal 3	--	Model Exam (75)	Unit I to V	CO1 to CO6
	5	MCQ (15)	Unit I to V	CO1 to CO6
	-	Course Attendance (10)	--	--

Submission Details:

Phase 1(Before AT 1)	Phase 2 (Before AT 2)	Phase 3 (Before Model)
Assignment 1	Assignment 2	Assignment 3

Google Class Code Details:
Class Name:

PLAN OF ASSESSMENT TEST –DISTRIBUTION OF MARKS:

TEST	CO- MARK WISE DISTRIBUTION						BLOOM'S LEVEL MARK WISE DISTRIBUTION					
AT-1	CO1	CO2	CO3	CO4	CO5	CO6	BTL1	BTL2	BTL3	BTL4	BTL5	BTL6
	37	23	-	-	-	-	37	23	-	-	-	-
AT-2	CO1	CO2	CO3	CO4	CO5	CO6	BTL1	BTL2	BTL3	BTL4	BTL5	BTL6
	-	-	37	23	-	-	22	22	16	-	-	-
MODEL	CO1	CO2	CO3	CO4	CO5	CO6	BTL1	BTL2	BTL3	BTL4	BTL5	BTL6
	20	20	20	20	10	10	50	30	20	-	-	-

Prepared By

Verified By

HOD/IT

Approved By

Principal