

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**

Name of the Faculty	Mrs.B.DEEPIKA			
Designation/Department	AP/CSE			
Course Code/Name	U20CS603/ DATA WAREHOUSING AND DATA MINING			
Year/Section/Department	III/A/CSE			
Credits Details	L:3	T: 0	P: 0	C:3
Total Contact Hours Required	45			

SYLLABUS:

UNIT I	INTRODUCTION TO DATA WAREHOUSING	No. of Periods: 9
Evolution of Decision Support Systems- Data warehousing Components –Building a Data warehouse, Data Warehouse and DBMS, Data marts, Metadata, Multidimensional data model, OLAP vs OLTP, OLAP operations, Data cubes, Schemas for Multidimensional Database: Stars, Snowflakes and Fact constellations.		
UNIT II	DATA WAREHOUSE PROCESS AND ARCHITECTURE	No. of Periods: 9
Types of OLAP servers, 3–Tier data warehouse architecture, distributed and virtual data warehouses. Data warehouse implementation, tuning and testing of data warehouse. Data Staging (ETL) Design and Development, data warehouse visualization, Data Warehouse Deployment, Maintenance, Growth, Business Intelligence Overview- Data Warehousing and Business Intelligence Trends - Business Applications- tools-SAS		
UNIT III	INTRODUCTION TO DATA MINING	No. of Periods: 9
Data mining-KDD versus data mining, Stages of the Data Mining Process-task primitives, Data Mining Techniques -Data mining knowledge representation – Data mining query languages, Integration of a Data Mining System with a Data Warehouse – Issues, Data preprocessing – 90 Data cleaning, Data transformation, Feature selection, Dimensionality reduction, Discretization and generating concept hierarchies-Mining frequent patterns- association-correlation		

UNIT IV CLASSIFICATION AND CLUSTERING	No. of Periods: 9
Decision Tree Induction - Bayesian Classification – Rule Based Classification –Classification by Back propagation – Support Vector Machines – Associative Classification – Lazy Learners – Other Classification Methods - Clustering techniques – , Partitioning methods- k-means Hierarchical Methods - distance-based agglomerative and divisible clustering, Density-Based Methods – expectation maximization -Grid Based Methods – Model-Based Clustering Methods – Constraint – Based Cluster Analysis – Outlier Analysis	
UNIT V DATA WAREHOUSING AND DATA MINING SOFTWARE’S AND APPLICATIONS	No. of Periods: 9
Mining complex data objects, Spatial databases, temporal databases, Multimedia databases, Time series and Sequence data; Text Mining –Graph mining-web mining-Application and trends in data mining	

Objective:

- This course deals with evolving multidimensional intelligent model from a typical system, representation of multidimensional data for a data warehouse.
- Discovering the knowledge imbibed in the high dimensional system, finding the hidden interesting patterns in data.
- Gives the idea to evaluate various mining techniques on complex data objects.

Text Book: (Details to be used in Reference Column of Course Plan)

1. Jiawei Han and MichelineKamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, third edition2011, ISBN: 1558604898. 2
2. Alex Berson and Stephen J. Smith, “Data Warehousing, Data Mining & OLAP”, TataMcGraw Hill Edition, Tenth Reprint 2007.
3. G. K. Gupta, “Introduction to Data Min Data Mining with Case Studies”, Easter EconomyEdition, Prentice Hall of India, 2006.

Reference Book: (Details to be used in Reference Column of Course Plan)

1. Mehmedkantardzic, “Dataminingconcepts,models,methods, and Algorithms”, Wiley Interscience, 2003.
2. Ian Witten, Eibe Frank, Data Mining; Practical Machine Learning Tools and Techniques, Third edition, Morgan Kaufmann, 2011.
3. George M Marakas, Modern Data Warehousing, Mining and Visualization, Prentice Hall, 2003.

Website: (Details to be used in Reference Column of Course Plan)

W1: <https://sisudata.com/blog/data-warehouses-decision-support-systems>.

W2: <https://www.guru99.com/oltp-vs-olap.html>

Online Mode of Study (if Any):

NPTEL/Course Era/Spoken Tutorial details can be listed.(Website link also to be listed)

❖ https://onlinecourses.nptel.ac.in/noc23_cs43/unit?unit=17&lesson=18

Course Plan:

Topic Number	Topic	Reference Detail	Page Number	Mode of teaching	Number of Periods Required	Cumulative Period
UNIT I - INTRODUCTION TO DATA WAREHOUSING						(9)
1	Evolution of Decision Support Systems	T1 T2 R3 W1	126 169-185 6	BB	1	1
2	Data warehousing Components	T1	127	BB	1	2
3	Building a Data warehouse	T2	115-127	BB	1	3
4	Data Warehouse and DBMS	T1 T2 R3	128-130 130-149 180-186	BB	1	4
5	Data marts, Metadata	T1	133-135	BB	1	5
6	Multidimensional data model	T2 R3	205-219 17-19	BB	1	6
7	OLAP vs OLTP	T1	135	BB	1	7
8	OLAP operations, Data cubes	T2	247-265	BB	1	8
9	Schemas for Multidimensional Database: Stars, Snowflakes and Fact constellations	T1	139	BB	1	9
Outcome of Unit I:						
CO1: Explain the basic concept of data warehousing.						
UNIT II - DATA WAREHOUSE PROCESS AND ARCHITECTURE						(9)
10	Types of OLAP servers, 3-Tier data warehouse architecture	T1 T2	131-133 251-255		1	10
11	Distributed and virtual data warehouses	T1	148-150	BB	1	11
12	Data warehouse implementation	T1	6.1	BB	1	12
13	Tuning and testing of data warehouse	T1	6.2	BB	1	13

14	Data Staging (ETL) Design and Development	T1	150-154	BB	1	14
15	Data warehouse visualization	T1	7.3	BB	1	15
16	Data Warehouse Deployment, Maintenance, Growth, Business Intelligence	T1	8.1-8.2	BB	1	16
17	Overview- Data Warehousing and Business Intelligence Trends	T1	9.1	BB	1	17
18	Business Applications- tools-SAS	T1	9.5-9.6	BB	1	18

Outcome of Unit II:

CO2: Understand the data warehouse process.

UNIT III - INTRODUCTION TO DATA MINING**(9)**

19	Data mining-KDD versus data mining	T3 T2 R3	4-9 331-336 81-83	BB	1	18
20	Stages of the Data Mining Process-task primitives	T1 T2	7-9 340-342	BB	1	20
21	Data Mining Techniques	T1	10-12	BB	1	21
22	Data mining knowledge representation	R3	78-80	BB	1	22
23	Data mining query languages, Integration of a Data Mining System with a Data Warehouse	T1	10.5	BB	1	23
24	Issues, Data preprocessing, Data cleaning	T1	84-86	BB	1	24
25	Data transformation, Feature selection,	T1	88-91	BB	1	25

	Dimensionality reduction					
26	Discretization and generating concept hierarchies	T1	111-118	BB	1	26
27	Mining frequent patterns, Association-correlation	T1	17-19	BB	1	27
Outcome of Unit III:						
CO3: Explain the concept of Data mining.						
UNIT IV - CLASSIFICATION AND CLUSTERING						(9)
28	Decision Tree Induction, Bayesian Classification	T1 T2	330-348 351-355	BB	1	28
29	Rule Based Classification, Classification by Back propagation	T1	355-359	BB	1	29
30	Support Vector Machines, Associative Classification	T1	408-412	BB	1	30
31	Lazy Learners, Other Classification Methods	T1	422	BB	1	31
32	Clustering techniques, Partitioning methods, K-means Hierarchical Methods	T1 T2	497-505 431-435	BB	1	32
33	Distance-based agglomerative and divisible clustering	T1	508-519	BB	1	33
34	Density-Based Methods, Expectation maximization, Grid Based Methods	T1	479-480	BB	1	34
35	Model-Based Clustering Methods, Constraint	T1	481	BB	1	35

36	Based Cluster Analysis, Outlier Analysis	T1	497	BB	1	36
Outcome of Unit IV: CO4: Illustrate classification and clustering techniques.						
UNIT V - DATA WAREHOUSING AND DATA MINING SOFTWARE'S AND APPLICATIONS (9)						
37	Mining complex data objects	T1	585-592	BB	1	37
38	Spatial databases	T1	593	BB	1	38
39	Temporal databases	T1	594-595	BB	1	39
40	Multimedia databases	R3	244	BB	1	40
41	Time series and Sequence data	R3	245-246	BB	1	41
42	Text Mining	R3	242	BB	1	42
43	Graph mining	R3	243	BB	1	43
44	Web mining	T3	272	BB	1	44
45	Application and trends in data mining	T1	607-615	BB	1	45
Outcome of Unit V: CO5: Explain the various databases. CO6: Explain the various mining.						

Course Outcome:

At the end of course:(**Consolidated outcome**)

Students should be able to do:

CO1: Explain the basic concept of data warehousing.

CO2: Understand the data warehouse process.

CO3: Explain the concept of Data mining.

CO4: Illustrate classification and clustering techniques.

CO5: Explain the various databases.

CO6: Explain the various mining.

Course Outcome Vs Program Outcome Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO 1	2	1	-	-	-	-	-	-	-	-	-	-	-	-
CO 2	2	1	-	-	-	-	-	-	-	-	-	-	-	-
CO 3	2	1	-	-	-	-	-	-	-	-	-	-	-	-
CO 4	2	1	-	-	-	-	-	-	-	-	-	-	-	-
CO 5	2	1	-	-	-	-	-	-	-	-	-	-	-	-
CO 6	2	1	-	-	-	-	-	-	-	-	-	-	-	-
AVG	2	1	-	-	-	-	-	-	-	-	-	-	2	-

Content Beyond Syllabus:

<ul style="list-style-type: none"> ❖ Big data ❖ R tool ❖ Data Analytics ❖ Weka tools ❖ Enterprise Data Warehouse

Web Portal	Assignment	Components	Topic Number with Topic/Unit Details
Web Portal 1	1	Assignment- Handwritten (20)	3.Building a Data warehouse /I 6.Multidimensional data model/I
	2	Poster/PPT Presentation (20)	13.Tuning and testing of data warehouse /II 15.Data warehouse visualization/II
Web Portal 2	3	Seminar(20)	19.Data mining-KDD versus data mining/III 27.Mining frequent patterns, Association-correlation/III
	4	Case Study Report/ Mini Project/ Model Making(20)	33.Distance-based agglomerative and divisible clustering/IV 36.Based Cluster Analysis, Outlier Analysis/IV
Web Portal 3	5	Technical Aptitude(15)	43.Graph mining/V 44.Text mining/V

		Attendance (Course attendance-10)	
--	--	---	--

Submission Details:

Phase 1(Before AT 1)		Phase 2 (Before AT 2)		Phase 3 (Before AT 3)
Assignment 1	Assignment 2	Assignment 3	Assignment 4	Assignment 5

Google Class Code Details: s374qmr

Prepared By

Mrs. B.DEEPIKA AP/CSE

Verified By

Dr.R.GOPI,HOD/CSE

Approved

Principal