



**BITS Pilani**  
Pilani | Dubai | Goa | Hyderabad

## BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI WORK INTEGRATED LEARNING PROGRAMMES

### COURSE HANDOUT

#### Part A: Content Design

<b>Course Title</b>	Advanced Data Mining
<b>Course No(s)</b>	SS ZG548
<b>Credit Units</b>	4
<b>Course Author</b>	Kamlesh Tiwari
<b>Version No</b>	
<b>Date</b>	

#### Course Objectives

No	Objective
CO1	To learn how to mine complex data (beyond conventional record data) and complex structures such as Tree/graph, sequence data, web/text data, stream data, mining multivariate time series data, high-dimensional data etc.
CO2	To learn how to apply these techniques to specific applications such as web search, Information Retrieval, social networks etc.
CO3	To learn about distributed computing solutions for data intensive applications in data mining

#### Text Book(s)

T1	David L. Olson & Dursun Delen “Advanced Data Mining”, Springer, 2008
----	--

#### Reference Book(s) & other resources

R1	Tan P. N., Steinbach M & Kumar V. “ <i>Introduction to Data Mining</i> ” Pearson Education, 2006
R2	Yates R. B. and Neto B. R. “ <i>Modern Information Retrieval</i> ” Pearson Education, 2005
R3	Han J. & Kamber M., “ <i>Data Mining: Concepts and Techniques</i> ”, Morgan Kaufmann Publishers, Second Edition, 2006
R4	Christopher D.M., Prabhakar R. & Hinrich S. “ <i>Introduction to Information Retrieval</i> ” Cambridge UP Online edition, 2009
R5	Hadzic F., Tan H. & Dillon T. S. “ <i>Mining data with Complex Structures</i> ” Springer, 20
R6	Agarwal Charu C. (Ed) “Data Streams Models and Algorithms” Springer 2007

## **Content Structure**

1. Introduction
  - 1.1. Review of data mining
  - 1.2. Objectives
  - 1.3. Overview
2. Incremental & Stream Data Mining
  - 2.1. Incremental Algorithms for Data Mining
  - 2.2. Characteristics of Streaming Data
  - 2.3. Issues and Challenges
  - 2.4. Streaming Data Mining Algorithms
3. Distributed computing solutions for data mining
  - 3.1. MapReduce/Hadoop
  - 3.2. Spark
4. Sequence Mining
  - 4.1. Characteristics of Sequence Data
  - 4.2. Problem Modeling
  - 4.3. Sequence Pattern Discovery
  - 4.4. Timing Constraints
5. Text Mining
  - 5.1. Text Classification
  - 5.2. Vector Space Model
  - 5.3. Flat and Hierarchical Clustering
  - 5.4. Streaming Data Mining Algorithms
6. Web Search
  - 6.1. Crawling & Indexing
  - 6.2. Hyperlink analysis
    - 6.2.1. HITS and Page Rank Algorithms
7. Mining Complex Structures
  - 7.1. Mining Trees
    - 7.1.1. Tree Miner
    - 7.1.2. Tree Model Guided Framework
    - 7.1.3. TMG framework for mining ordered & unordered subtrees
  - 7.2. Mining Graphs
    - 7.2.1. Approaches to graph mining
  - 7.3. Case Study: Information Retrieval
  - 7.4. Case Study: Mining Social Networks

## **Learning Outcomes:**

No	Learning Outcomes
LO1	To understand how to update the patterns incrementally when the data is continuously coming
LO2	To understand the role of distributed computing in data intensive data mining
LO3	To study how to investigate the sequence data
LO4	To understand how text mining is different from data mining and how to mine it
LO5	To understand what goes into the web search and to study methods of web search and their improvements
LO6	To understand how to mine complex structures other than records while retaining the relations

	among the entities
--	--------------------

## Part B: Contact Session Plan

<b>Academic Term</b>	First Semester 2021-2022
<b>Course Title</b>	Advanced Data Mining
<b>Course No</b>	SS ZG548
<b>Lead Instructor</b>	Kamlesh Tiwari

### Course Contents

<b>Contact Hours(#)</b>	<b>List of Topic Title (from content structure in Course Handout)</b>	<b>Text/Ref Book/external resource</b>
1	Introduction Review and Overview	
2	Incremental Data Mining Relook traditional algorithms	See Class Slides
3, 4, 5	Incremental algorithms and their design and analysis	See Class Slides
6	Stream Data Mining Characteristics, Issues and Challenges	R6 Ch1,4 See Class Slides
7, 8	Stream Data Mining Algorithms and their Comparison	R6 Ch1,4 See Class Slides
9, 10, 11, 12	Distributed computing solutions for data mining	See Class Slides
13	Sequence Mining Characteristics and Problem Modeling	R1 7.4
14	Sequence Pattern Discovery Timing Constraints	R1 7.4
15	Text Mining Data Representation and Characteristics	R4 Ch 1, 13, R2 Ch 7
16	Text Classification Feature Selection & Models	R4 Ch 14, R2 Ch 7
17	Text Classification Vector Space Model	R4 Ch 14, R2 Ch 7
18	Text Classification Multiclass classifiers for text	R4 Ch 13, 14
19	Text Clustering Flat and hierarchical	R4 Ch 16,17

20	Web Search	R4 Ch 1, 6, 19
21, 22, 23	Crawling & Indexing	R4 Ch 20
24	Link Analysis	R4 Ch 21 See Class slides
25	Mining Complex Structures Data Representation	R5 Ch1 See Class slides
26	Tree Mining problem and Tree basics	R5 Ch 2, 3 See Class slides
27	Tree Miner	R5 Ch 3 See Class slides
28	TMG Model Guided Framework	R5 Ch 4, 5, 6
29	Graph Mining Introduction and applications	R5 Ch 11 See Class slides
30	Case Study: Information Retrieval	See Class slides
31	Case Study: Social Network Mining	See Class slides
32	Case Study: Social Network Mining	See Class slides

### Evaluation Scheme:

Legend: EC = Evaluation Component; AN = After Noon Session; FN = Fore Noon Session

No	Name	Type	Duration	Weight	Day, Date, Session, Time
EC-1	Quiz-I/ Assignment-I	Online	-	5%	August 16-30, 2021
	Quiz-II			5%	September 16-30, 2021
	Quiz-III/ Assignment-II			5%	October 16-30, 2021
EC-2	Mid-Semester Test	Open Book	2 hours	35%	Friday, 24/09/2021 (FN) 10 AM - 12 Noon
EC-3	Comprehensive Exam	Open Book	2 hours	50%	Friday, 12/11/2021 (FN) 10 AM - 12 Noon

**Note** - Evaluation components can be tailored depending on the proposed model.

### Important Information

Syllabus for Mid-Semester Test (Open Book): Topics in Weeks 1-7

Syllabus for Comprehensive Exam (Open Book): All topics given in plan of study

### Evaluation Guidelines:

1. EC-1 consists of either two Assignments or three Quizzes. Announcements regarding the same will be made in a timely manner.
2. For Closed Book tests: No books or reference material of any kind will be permitted. Laptops/ Mobiles of any kind are not allowed. Exchange of any material is not allowed.
3. For Open Book exams: Use of prescribed and reference text books, in original (not

photocopies) is permitted. Class notes/slides as reference material in filed or bound form is permitted. However, loose sheets of paper will not be allowed. Use of calculators is permitted in all exams. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.

4. If a student is unable to appear for the Regular Test/Exam due to genuine exigencies, the student should follow the procedure to apply for the Make-Up Test/Exam. The genuineness of the reason for absence in the Regular Exam shall be assessed prior to giving permission to appear for the Make-up Exam. Make-Up Test/Exam will be conducted only at selected exam centres on the dates to be announced later.

It shall be the responsibility of the individual student to be regular in maintaining the self-study schedule as given in the course handout, attend the lectures, and take all the prescribed evaluation components such as Assignment/Quiz, Mid-Semester Test and Comprehensive Exam according to the evaluation scheme provided in the handout.