BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, Pilani
Pilani Campus

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI
INSTRUCTION DIVISION
Second Semester 2015-2016
COURSE HANDOUT (PART II)

In addition to Part-I (general handout for all courses appended to this time table) this portion gives further details pertaining to the course.

Course No.: CS G520/SS G520
Course Title: **Advanced Data Mining**
Instructor-in-charge: **POONAM GOYAL (poonam@pilani.bits-pilani.ac.in)**

1. Objective and Scope

Advanced Data Mining is a special topic course on Data Mining. Topics covered go beyond conventional record data mining to mining complex data structures and complex data: Tree/graph, sequence data, web/text data, stream data, spatiotemporal data, mining multivariate time series data, high-dimensional data. The course also deals with mining social networking sites, data in multiple relations (Multi-relational Data Mining) and with distributed computing solutions for data intensive applications.

2. Text Book

David L. Olson & Dursun Delen ***"Advanced Data Mining",*** Springer, 2008.

3. Reference Books

R1: Hadzic F., Tan H. & Dillon T. S. "***Mining data with Complex Structures***" Springer, 2011
R2: Yates R. B. and Neto B. R. "***Modern Information Retrieval***" Pearson Education, 2005
R3: Tan P. N., Steinbach M & Kumar V. "***Introduction to Data Mining"*** Pearson Education, 2006
R4: Han J. & Kamber M., "***Data Mining: Concepts and Techniques",*** Morgan Kaufmann Publishers, Second Edition, 2006
R5: Christopher D.M., Prabhakar R. & Hinrich S. "***Introduction to Information Retrieval***" Cambridge UP Online edition, 2009

4. Course Plan

Lecture No.	Learning Objective	Topic(s)	Chapter Reference
1-2	To understand the objectives of the course	**Introduction & Review**	Class Notes
3-6	To understand how to update the patterns incrementally when the data is continuously coming	**Incremental & Stream Data Mining** • Incremental Algorithms for Data Mining • Characteristics of Streaming Data • Issues and Challenges • Streaming Data Mining Algorithms	Class Notes
7-9	To understand the role of distributed computing in data intensive data mining	**Distributed computing solutions for data mining** • MapReduce/Hadoop • Cluster Computing	Class Notes
10-15	To understand how to mine complex structures other than records	**Mining Complex Structures** • Algorithmic Development Issues • Mining trees ○ Tree Model Guided Framework ○ TMG framework for mining ordered & unordered subtrees ○ Mining distance-constrained embedded subtrees ○ Tree Mining Applications	R1: 2,5,6,8 Class Notes

		• Mining Graphs ○ Approaches to graph mining ▪ Apripori like Methods ▪ Pattern Growth methods ▪ Greedy Search Methods	
16-17	To understand the application of data mining in Social Networks	**Case study:** Mining Social Networks	R5:7 Class Notes
18-20	To study how to investigate the sequence data	**Sequence Mining** • Characteristics of Sequence Data • Problem Modeling • Sequential Pattern Discovery • Timing Constraints • Applications in Bioinformatics	R3: 7.4
21-24	To understand how text mining is different from data mining and how to mine it	**Text Mining** • Text Classification • Vector Space Model • Flat and Hierarchical Clustering	R2:7 R5:13,14, 16,17
25-32	To understand what goes into the web search and to study methods of web search and their improvements	**Web Search** • Crawling & Indexing • Hyperlink Analysis • Page Rank algorithm • Web Search and Information Retrieval • Case Study: Query Recommender System	R2: 8,12,13 R5: 20,21
33-35	To understand the characteristics of MVTS data and the need for mining MVTS data	**Multivariate Time Series (MVTS) Mining** • Importance of MVTS data • Sources of MVTS data • Mining MVTS data ○ Sign Language Data ○ Agro-meteorological Data	Class Notes
36-40	To understand how to mine in several relational tables exploiting the relation among them	**Multi-relational Data Mining (MRDM)** • Introduction • Relational Patterns • Applications • Inductive Logic Programming	Class Notes

5. Evaluation Schedule

Component	Duration	Weightage(%)	Date & Time	Venue	Remarks
Test	90 Mins.	25	18/3 4:00- 5:30 PM		Open Book
Labs/Assignments		40			
Comprehensive	3 Hours	35	13/5 AN		Partly open

7. Labs/ Assignments

Labs/assignments will be given to the students to understand the topics covered in the class. These assignments will immensely help the students in gaining a better understanding of recent developments in the subject and how to apply the concepts of data mining on to various applications.

9. Chamber Consultation Hours

To be announced in the class.

10. Make-up Policy: Prior Permission is must and Make-up shall be granted only in genuine cases based on individual's need and circumstances.

11. Notices

All the notices concerning this course will be displayed on the **CSIS Notice Board** or **course website**.

12. Reading Material

Research papers and other reading material will be provided on the course website.

Instructor-in-charge