BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, Pilani
Pilani Campus

INSTRUCTION DIVISION
First Semester 2016-2017
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 F415
Course Title: **Data Mining**
Instructor-in-charge: **POONAM GOYAL (poonam@bits-pilani.ac.in)**

1. Objective and Scope

The course explores the concepts and techniques of data mining, a promising and flourishing frontier in database systems. Data Mining is automated extraction of patterns representing knowledge implicitly stored in large databases, data warehouses, and other massive information repositories. It is a decision support tool that addresses unique decision support problems that cannot be solved by other data analysis tools such as Online Analytical Processing (OLAP). The course covers data mining tasks like constructing decision trees, finding association rules, classification, and clustering. The course is designed to provide students with a broad understanding in the design and use of data mining algorithms. The course also aims at providing a holistic view of data mining. It will have database, statistical, algorithmic and application perspectives of data mining.

2. Text Book

i) Tan P. N., Steinbach M & Kumar V. "***Introduction to Data Mining"*** Pearson Education, 2006.

3. Reference Books

i) Han J & Kamber M, "***Data Mining: Concepts and Techniques",*** Morgan Kaufmann Publishers, Second Edition, 2006
ii) Zaki MJ & Wagner M JR, "***Data Mining and Analysis-Fundamental Concepts and Algorithms***" Cameridge Univ Press, 2014.
iii) Dunhum M.H. & Sridhar S. "***Data Mining-Introductory and Advanced Topics***", Pearson Education, 2006.

4. Course Plan

Lecture No.	Learning Objective	Topic(s)	Chapter Reference
1-2	To understand the definition and applications of Data Mining	**Introduction to Data Mining** • Motivation • What is Data Mining? • Data Mining Tasks • Issues in Data Mining • Applications	1+Class Notes
3-5	To understand types of data and to improve the quality of data and efficiency and the ease of the mining process.	**Data Preprocessing** • Types of data • Data Quality • Data preprocessing • Similarity and Dissimilarity Measures	2
6	To study how to investigate the data	**Data Exploration** • Data Set & its Statistics • Visualization • OLAP & Multidimensional Data Analysis	3 Self Study
7-10	To understand Classification and its applications	**Classification** • Introduction • Applications • Decision Tree based Algorithms	4+Class Notes

		• Model Over-fitting • Performance Evaluation of a Classifier • Comparing Classifiers	
11-15	To study the alternative approaches for Classification	**Classification: Alternative Techniques** • Rule Based Classifier • Nearest Neighbor Classifier • Bayesian Classification • Support Vector Machine • Ensemble Classifiers • Class Imbalance Problem • Multiclass Problem	5
16-19	To understand applications of Association Rule Mining and algorithms to find them	**Association Rule Mining** • Introduction • Applications • Market-Basket Analysis • Frequent Itemsets • Apriori Algorithm • Alternative Methods	6
20-23	To understand methods and need for finding complex Association Rules	**Advanced Association Rule Mining** • Generalized Association Rules • Multilevel Association Rules • Multidimensional Association Rules • Temporal Association Rules • Infrequent Patterns • Constrained Based Association Rules	7+Class Notes
24-28	To understand applications and algorithms for Clustering	**Clustering** • Introduction • Applications • Partitioning Algorithms • Hierarchical Algorithms • Density based Algorithms • Cluster Evaluation	8
29-33	To study advanced topics in cluster analysis	**Clustering: Additional Issues and Algorithms** • Characteristics of Data, Clusters and clustering Algorithms • Graph Based Clustering • Scalable Clustering Algorithms	9
34-35	To understand detection of anomalies & their causes	**Anomaly Detection** • Preliminaries • Statistical Approaches • Proximity based Outlier Detection • Density based Outlier Detection • Clustering Based Techniques	10
36-40	To introduce advanced topics in Data Mining	**Advanced Topics** • Web Mining • Incremental Algorithms for Data Mining • Stream Data Mining	Class Notes

5. Evaluation Schedule

Component	Duration	Weightage(%)	Date & Time	Venue	Remarks
Mid Sem Exam	90 Mins.	30			Closed Book

Labs/Assignments/quiz(s)		30	To be announced		
Comprehensive	3 Hours	40	07/05		Partly open

7. **Labs**
   Two hour lab will be conducted every week. Students will be applying the concepts of data mining on the problems and cases through the Data Mining software, IBM SPSS Modeler. Students will also be exposed to modeling of the problems.
8. **Assignments**
   Assignment(s) (programming/reading) will be given to the students. This will immensely help the students in gaining a better understanding of the subject.
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**.

Instructor-in-charge