BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

HYDERABAD CAMPUS

**Second Semester 2013 – 2014**

**INSTRUCTION DIVISION**

**COURSE HANDOUT (PART II)**

**Date**: 6th January 2014

In addition to Part-I (General Handout for all courses appended to this time table) this portion gives specific details regarding the course.

**Course No : CS C415**

**Course Title : Data Mining**

**Instructor-in-charge : Dr. N.L.Bhanu Murthy**

1. **Course Description**

In this course we explore how this interdisciplinary field brings together techniques from databases, statistics, machine learning, and information retrieval. We will discuss the main data mining methods currently used, including data warehousing and data cleaning, clustering, classification, association rules mining, and web mining. Designing algorithms for these tasks is difficult because the input data sets are very large, and the tasks may be very complex. One of the main focuses in the field is the integration of these algorithms with relational databases and we will examine the additional complications.

1. **Course** **Pre-requisites**

###### An upper-level undergraduate course(s) in algorithms and data structures, a basic course on probability and statistics. A DBMS course is helpful but not necessary.

###### **Course** **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.

1. **Text Book**

T1. Tan,Pang-Ning & others. “***Introduction to Data Mining”*** Pearson Education, 2006.

1. **Reference Books**

R1. Han J & Kamber M, “***Data Mining: Concepts and Techniques”,*** Morgan Kaufmann Publishers, Second Edition, 2006

R2. Luis Torgo, “Data Mining with R Leaning with case studies”, CRC Press.

R3. Dunhum M.H. & Sridhar S. “***Data Mining-Introductory and Advanced Topics***”, Pearson Education, 2006.

R4. Grigoris Antoniou and Frank van Harmelen “**A Semantic Web Primer**”, The MIT Press Cambridge, Massachusetts London, England 2003

R5. S.Sumathi & S.N.Sivanandam “**Introduction to Data mining and its applications**”, Springer-verlag

1. **Lecture Schedule**

|  |  |  |  |
| --- | --- | --- | --- |
| **Lecture No.** | **Learning Objective** | **Topic(s)** | **Chapter Reference** |
| 1 | To understand the definition and applications of Data Mining | Introduction to Data Mining  * Motivation * What is Data Mining? * Data Mining Tasks | 1+Class Notes |
| 2-3 | * Issues in Data Mining  Applications | 1+Class Notes |
| 4-5 | 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 | 2 |
| 6-7 | * Data preprocessing * Similarity and Dissimilarity | 2 |
| 8-9 | To study how to investigate the data | **Data Exploration**   * Data Set & its Statistics * Visualization | 3 |
| 9-11 | * OLAP & Multidimensional Data Analysis | 3 |
| 12 | To understand applications of Association Rule Mining and algorithms to find them | Association Rule Mining  * Introduction * Applications * Market-Basket Analysis | 6 |
| 13-15 | * Frequent Itemsets * Apriori Algorithm | 6 |
| 16-17 | * Alternative Methods | 6 |
| 18 | To understand methods and need for finding complex Association Rules | Advanced Association Rule Mining  * Generalized Association Rules * Multilevel Association Rules | 7 |
| 19-22 | * Multidimensional Association Rules * Infrequent Patterns | 7 |
| 23 | * Constrained Based Association Rules | 7 |
| 24-26 | To understand applications and algorithms for Clustering | Clustering  * Introduction * Applications * Partitioning Algorithms | 8 |
| 27-29 | * Hierarchical Algorithms * Density based Algorithms | 8 |
| 30 | * Cluster Evaluation | 8 |
| 31-32 | To study advanced topics in cluster analysis | Anomaly Detection  * Preliminaries * Statistical Approaches | 10 |
| 33-34 | * Proximity based Outlier Detection * Density based Outlier Detection | 10 |
| 35 | Clustering Based Techniques | 10 |
| 36 |  | * Mining the World Wide Web | 10 |
| 37 - 40 |  | Advanced Classification Techniques | Class Notes |

**Evaluation Schedule**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Component | **Mode** | **Duration** | **Date & Time** | **Weightage(%)** |
| Test-I | Closed Book | 60 Mins. | 24/02/2014, 2-3 pm | 20 |
| Test-II | Closed Book | 60 Mins. | 28/03/2014, 2-3 pm | 20 |
| Assignments | Open Book | -- |  | 20 |
| Comprehensive | Closed Book | 3 Hours | 16/05/2014 FN | 40 |

**7. Labs/Assignments**

A series of programming and reading assignments will be given to the students on a regular basis. These assignments will immensely help the students in gaining a better understanding of the subject.

1. **Make-up Policy:**  Prior Permission is mustand Make-up shall be granted only in genuine cases based on individual’s need and circumstances.
2. **Course Notices**

All notices pertaining to this course will be displayed on the CSIS Notice Board.

1. **Chamber Consultation Hours**

Thursday 4PM – 5PM @B219

**Instructor-in-charge**

**CSC415**