BITS PILANI, DUBAI CAMPUS
ACADEMIC - UNDERGRADUATE STUDIES DIVISION
FIRST SEMESTER 2023 – 2024

Course Handout (Part – II)

Date: 28.08.2023

In addition to Part I (General Handout for all courses appended to the Time Table) this portion further specific details regarding the course.

Course No. : **CS F415 (3 0 3)**
Course Title : **Data Mining**
Instructor-in-charge : **Dr. J Angel Arul Jothi**
Instructor : **Dr. J Angel Arul Jothi**

Scope and Objective of the Course:

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. Students will also get to apply the DM concepts using Python.

Course Pre/Co- requisite (if any) **& Catalogue/Bulletin Description:**

Given in the Bulletin 2023–2024

Text book(s):

TB1. Tan P. N., Steinbach M & Kumar V. "***Introduction to Data Mining"*** Pearson Education, 2ed, 2021.

Reference book(s):

RB1. Han J & Kamber M, **"*Data Mining: Concepts and Techniques",*** Morgan Kaufmann Publishers, 2012

RB2. Dunhum M.H. & Sridhar S. "***Data Mining-Introductory and Advanced Topics***", Pearson Education, 2006.

Course Plan:

Lec. No	Learning objectives	Contents	Text/Reference Book
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	Class Notes TB1: Ch1 RB1: Ch1
3-6	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	TB1: Ch 2 RB1: Ch 3
7	To study how to investigate the data	**Data Exploration** ● Data Set & its Statistics ● Visualization	TB1: Ch 3
8-12	To understand Classification and its applications	**Classification** ● Introduction ● Applications ● Decision Tree based Algorithms ● Model Overfitting ● Performance Evaluation of a Classifier ● Comparing Classifiers	TB1: Ch 4 RB1: Ch 7
13 – 19	To study the alternative approaches for Classification	**Classification: Alternative Techniques** ● Rule Based Classifier ● Nearest Neighbor Classifier	TB1: Ch 5 RB1: Ch 7

		● Bayesian Classification ● Support Vector Machine ● Ensemble Classifiers ● Class Imbalance ● Multiclass Problem	
20 – 25	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	TB1: Ch 6 RB1: Ch 6
26 – 27	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	Class Notes RB1: Ch 5
28 – 33	To understand applications and algorithms for Clustering	**Clustering** ● Introduction ● Applications ● Partitioning Algorithms ● Hierarchical Algorithms ● Density based Algorithms ● Cluster Evaluation	TB1: Ch 8 RB1: Ch 8
34 – 36	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	TB1: Ch 9 RB1: Ch 8
37 – 39	To understand detection of anomalies & their causes	**Anomaly Detection** ● Preliminaries ● Statistical Approaches ● Proximity based Outlier Detection ● Density based Outlier Detection ● Clustering Based Techniques	TB1: Ch10
40 - 42	To introduce advanced topics in Data Mining	**Advanced Topics** ● Artificial Neural Networks ● CNN and its applications	TB1: Ch 5 Class Notes

* The lectures may be slightly diverging from aforesaid plan based on students' background & interest in the topic, which may perhaps include special lectures and discussions that would be planned and schedule notified accordingly

Course Learning Outcomes (CLOs)

Upon successful completion of this course, students should be able to:

- **CLO1** Differentiate different data mining tasks and explain various data mining algorithms
- **CLO2** Perform the data preprocessing and understand the implications
- **CLO3** Solve numerical problems using data mining techniques on a given dataset
- **CLO4** Evaluate data mining algorithms using suitable performance metrics
- **CLO5** Identify real world problems, apply state-of-the-art data mining algorithms, compare different approaches, discuss the findings and prepare a report

Evaluation scheme:

EC No.	Components	Nature of the Component	Duration	Weightage %	Date & Time	Venue
1	Project Assignment 1**	Open book	Continuous	10%	On or before 29.09.2023	To be announced
2	Mid Sem Test	Closed Book	90 mins	30%	31.10.2023 FN	
3	Project Assignment 2**	Open book	Continuous	20%	On or before 30.11.23	
4	Compre Exam	Closed Book	3 hours	40%	05.01.2024 FN	

**Detailed instructions will be communicated to the students.

**Student project demonstration and presentation will be conducted

Mapping of CLOs, PLOs, and CECs

CLOs	PLOs	Evaluation Components (ECs)			
		EC1	EC2	EC3	EC4
CLO1	1, 2, 4		✓		✓
CLO2	1, 3, 4		✓		✓
CLO3	1, 3, 4		✓		✓
CLO4	1, 3, 4		✓		✓
CLO5	1, 3, 4, 8	✓		✓	

Mid-sem Grading:
Mid-sem grading will be displayed after two evaluation components or earlier whenever about 40% of evaluation components are completed.
Note: A student will be likely to get "NC", if he / she doesn't appear / appear for the sake of appearing for the evaluation components / scoring zero in pre-compre total.

Makeup and Attendance policies:
Make-ups are not given as a routine. It is solely dependent upon the genuineness of the circumstances under which a student fails to appear in a scheduled evaluation component. In such circumstances, prior permission should be obtained from the Instructor-in-Charge (I/C). **Students with less than 50% of attendance will not be allowed to avail the make-ups.** The decision of the I/C in the above matter will be final.

Attendance: Every student is expected to be responsible for regularity of his/her attendance in class rooms and laboratories, to appear in scheduled tests and examinations and fulfill all other tasks assigned to him/her in every course. A student should have a minimum of 60% of attendance in a course to be eligible to appear for the Comprehensive Examination in that course. For the students under the purview of Academic Counseling Board (ACB), the Board shall prescribe the minimum attendance requirement on a case-to-case basis. Attendance in the course will be a deciding factor in judging the seriousness of a student which may be directly / indirectly related to grading.

General timings for consultation:
Chamber consultation hour: Friday 5th hour

General instructions:
Students should come prepared for classes and carry the text book(s) or material(s) as prescribed by the Course Faculty to the class.

Notices:
All notices concerning the course will be displayed on the CS department Notice Board and on LMS Moodle. Students are also informed to check their BITS email (regularly) for any communication regarding the course.

Dr. J. Angel Arul Jothi
Instructor-in-Charge

Instructor's Contact Details:
Dr. J. Angel Arul Jothi, Room No: 231, Contact No: +9714-4200700 Ext.158,
e-mail: angeljothi@dubai.bits-pilani.ac.in