Birla Institute of Technology & Science, Pilani
Hyderabad Campus

SECOND SEMESTER 2022-2023

Course Handout Part II

Date: 10th Jan 2023

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

Course No. : *CS F415*
Course Title : Data Mining
Instructor-in-Charge : Dr. Manik Gupta (*manik@hyderabad.bits-pilani.ac.in*)

Scope and Objective of the Course:

The course explores the concepts and techniques of data mining, a promising and flourishing frontier in data science. Analyzing large amounts of data has become a necessity and the problem therefore arises as to "how to analyze the data". Data Mining is the subject dealing automated extraction of useful information or patterns representing knowledge implicitly stored in large databases, data warehouses, and other massive data repositories. It is a technology that blends traditional data analysis methods with sophisticated algorithms for association rule mining, clustering, classification and outlier analysis. The course is designed to provide students with a broad understanding in the design and use of data mining algorithms. The course will provide an algorithmic as well as application perspectives of data mining.
At the end of the course the student should be able to

- Choose an appropriate data preprocessing techniques based on the given data.
- Identify and design an appropriate data mining task and analysis technique given a problem.
- Gain practical hands on experience in implementing data mining algorithms.

Textbooks:

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

Reference Books:

R1. Han J & Kamber M, "***Data Mining: Concepts and Techniques,"*** Morgan Kaufmann Publishers, Second Edition, 2006
R2. Christopher Bishop**: *"Pattern Recognition and Machine Learning"***, Springer International Edition
R3. Tom M. Mitchell: "***Machine Learning***", The McGraw-Hill Companies, Inc.
R4. Charu C. Aggarwal "***Outlier Analysis***" Springer International Publishing (2017)

Course Plan:

Lecture No.	Learning objectives	Topics to be covered	Chapter in the Text Book
1-2	To be able to define and list applications of Data Mining	**Introduction to Data Mining** ● Motivation ● What is Data Mining? ● Data Mining Tasks ● Issues in Data Mining ● Applications	T1.1
3-8	• To be able to list preprocessing steps and identify right preprocessing step given the data • To be able to perform dimensionality reduction on huge data using PCA and feature selection approaches	**Data Preprocessing** ● Types of data ● Data Quality ● Data preprocessing ● Similarity and Dissimilarity **Dimension Reduction** ● Principal Component Analysis ● Greedy Algorithms for feature selection	T1.2
9-14	To be able to apply and implement association rule mining	**Association Rule Mining** ● Introduction ● Applications ● Market-Basket Analysis ● Frequent Itemsets ● Apriori Algorithm ● FP Growth	T1.6
15-18		**Advanced Association Rule Mining** ● Graph Mining ● Sequence Mining	T1.7
19-28	To be able to apply and implement unsupervised learning algorithms	**Clustering** ● Introduction	T1.8 T1.9

		● Applications ● Partitioning Algorithms ● Hierarchical Algorithms ● Density based Algorithms ● Cluster Evaluation ● Graph-Based Clustering	
29-32	To be able to apply and implement anomaly detection algorithms	**Outlier Analysis** ● What are Outliers ● Distance-Based Outlier Analysis ● Density-Based Outliers	T1.10
33-38	To be able to apply and implement classification models	**Classification Techniques** ● Basic Classification Techniques ● Decision Tree ● Naïve Bayes	T1.4 T1.5
39-40	Buffer	**Review**	

Evaluation Scheme:

Component	Duration	Weightage (%)	Date & Time	Nature of Component
Mid Term Exam	1.5 hours	30	14/03 11.30 - 1.00PM	Closed Book
Project(at least 10% before Mid sem)	--	30	TBA	Take home
Comprehensive Exam	3 Hours	40	10/05 AN	Closed Book

Chamber Consultation Hour:

H-126, Wed-Friday 11am to 12pm

Notices:

All notices and announcements pertaining to this course will be displayed on the CMS/Google Classroom.

Make-up Policy:

1. No Make-up requests for project submissions will be catered to.

2. Prior permission of the Instructor-in-Charge is required to get make-up for the mid-semester exam. Only on producing documentary proof of absence minimum one day prior to the exam, proving that student would be physically unable to appear for the exam, the decision of granting the make-up will be taken. The recommendation from chief warden is necessary to request for a make-up.
3. Prior permission of Dean, AUGSD is required to get make-up for the comprehensive exam.
4. Instructor-in-charge's/Dean's decision in the matter of granting make-up would be final.

Academic Honesty and Integrity Policy:

Academic honesty and integrity are to be maintained by all the students throughout the semester and no type of academic dishonesty is acceptable.

INSTRUCTOR-IN-CHARGE
CS F415