Birla Institute of Technology & Science, Pilani
Hyderabad Campus

SECOND SEMESTER 2021-2022

Course Handout Part II

Date: 15-01-2022

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 Apurba Das

Scope and Objective of the Course:

The course explores the concepts and techniques of data mining, a promising and flourishing frontier in database science. 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 finding association rules, classification, and clustering techniques. 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.

At the end of the course the student should be able to

- Choose appropriate data preprocessing techniques based on the given data.
- Compare the performance of binary and Multi class classification algorithms and fine tune the parameters of the algorithms.
- Write efficient implementations of data mining algorithms.
- Select appropriate clustering techniques.
- Design appropriate data mining technique given a problem.

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..

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	T1.1 + Class Notes
		• Issues in Data Mining • Applications	
3-4	To be able to list preprocessing steps and identify right preprocessing step given the data	**Data Preprocessing** • Types of data • Data Quality • Data preprocessing • Similarity and Dissimilarity	T1.2
5-8	To be able to perform dimension reduction on huge data using PCA and feature selection approaches	**Dimension Reduction** • Principal Component Analysis • Greedy Algorithms for feature selection	T1.2 R2.12
9-15	To be able to apply and implement classification models	**Classification Techniques** • Basic Classification Techniques • SVM	T1.5 R2.9
16-22	To be able to apply and implement association rule mining	**Association Rule Mining** • Introduction • Applications • Market-Basket Analysis • Frequent Itemsets • Apriori Algorithm • Alternative Methods	T1.6
23-31		**Advanced Association Rule Mining** • Generalized Association Rules • Multilevel Association Rules • Graph Mining • Sequence Mining • Multidimensional Association Rules • Constrained Based Association Rules	T1.7
32-40	To be able to apply and implement unsupervised learning algorithms	**Clustering** • Introduction • Applications • Partitioning Algorithms • Hierarchical Algorithms • Density based Algorithms • Cluster Evaluation • EM Algorithm • Graph-Based Clustering	T1.8 T1.9 R2.12 R3.6

Evaluation Scheme:

Component	Duration	Weightage (%)	Date & Time	Nature of Component
Mid Term Exam	90 Mins.	25	11/03 3.30pm to5.00pm	Closed Book
Class Project	--	15	TBA	Open Book
Assignments (programming+theory)	--	20	TBA	Open Book
Comprehensive	120 min	40	10/05 AN	Closed Book

Note: 40% of the evaluation to be completed by midsem grading.
"For Comprehensive exam and Mid-semester Test, the mode (offline/online) and the duration are subject to changes as decided by the AUGSD/Timetable division in future."

Chamber Consultation Hour: To be announced in the class.

Notices: All notices pertaining to this course will be announced in the google-classroom.

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

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