BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
WORK INTEGRATED LEARNING PROGRAMMES

COURSE HANDOUT

Part A: Content Design

Course Title	Data Mining
Course No(s)	DSECLZC415
Credit Units	3
Revision Date	06/11/2020

Course Description

Data Mining is automated extraction of patterns representing knowledge implicitly stored in information repositories. The course covers how to prepare real-world data for data mining tasks and perform data mining tasks such as classification, association, and clustering. Students gain knowledge of the design and use of data mining algorithms. The course includes statistical, algorithmic and application perspectives of data mining.

Course Objectives

CO1	Understand the importance of data mining and the knowledge discovery that can be made from information repositories with the help of data mining
CO2	Understand techniques of preparing real-world data for performing data mining
CO3	Understand data mining techniques for discovering interesting patterns from data
CO4	Understand efficiency, effectiveness of applicable techniques for data mining.

Text Book(s)

T1	Tan P. N., Steinbach M, Karpatne A, & Kumar V. “Introduction to Data Mining” Pearson Education, 2019
T2	Data Mining: Concepts and Techniques, Third Edition by Jiawei Han and Micheline Kamber Morgan Kaufmann Publishers, 2011

Reference Book(s) & other resources

R1	Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner by Vijay Kotu and Bala Deshpande Morgan Kaufmann Publishers © 2015
R2	Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications by Gary Miner et al. Academic Press © 2012
R3	Recommender Systems for Learning by Nikos Manouselis, Hendrik Drachsler, Katrien Verbert and Erik Duval Springer © 2013
R4	Mining of Massive Datasets 3rd ed by Jure Leskovec, Anand Rajaraman, Jeffrey Ullman
R5	https://www.sciencedirect.com/science/article/pii/S2212567115014859

Modular Content Structure

1. **Introduction to Data Mining**
   - 1.1. Data Mining definitions
   - 1.2. Data Mining activities
   - 1.3. DM process
   - 1.4. DM challenges
2. **Data Preprocessing**
   - 2.1. Data Quality
   - 2.2. Data preprocessing requirements
   - 2.3. Data preprocessing techniques
3. **Data Exploration**
   - 3.1. Statistical descriptions of data
   - 3.2. Measuring data similarity & dissimilarity
4. **Classification and Prediction**
   - 4.1. Concepts of classification and prediction
   - 4.2. Decision trees for classification
   - 4.3. Rule based classification,
   - 4.4. Prediction Techniques
5. **Association Analysis**
   - 5.1. Association analysis concepts
   - 5.2. Apriori Algorithm for frequent itemsets
   - 5.3. FP-Tree technique for frequent itemsets
   - 5.4. Mining association rules
6. **Clustering**
   - 6.1. Cluster analysis concepts.
   - 6.2. Partitioning methods
   - 6.3. Hierarchical methods for cluster analysis
   - 6.4. Density based methods for cluster analysis
7. **Anomaly Detection**
   - 7.1. Concepts of Outliers
   - 7.2. Statistical approaches
   - 7.3. Proximity and Density based outlier detection
8. **Data mining on unstructured (Big) data**
   - 8.1. Graph Mining methods and applications
   - 8.2. Multimedia Data Mining
   - 8.3. Text Mining, Web and Social Media Mining
9. **Data Mining Applications**
   - 9.1. Recommendation systems
   - 9.2. Fraud Detection
   - 9.3. Sentiment Analysis

Learning Outcomes:

No	Learning Outcomes
LO1	Realize how data mining can enable knowledge discovery.
LO2	Knowledge of techniques of preparing real-world data for performing data mining.
LO3	Knowledge of data mining techniques for discovering interesting patterns from data.
LO4	Knowledge on efficiency, effectiveness of applicable techniques for data mining.

Part B: Contact Session Plan

Academic Term	S2 2021-22
Course Title	Data Mining
Course No	DSECLZC415
Lead Instructor	T V Rao

Course Contents

Contact Hours(#)	List of Topic Title (from content structure in Part A)	Topic # (from content structure in Part A)	Text/Ref Book/external resource
1 2	• Introduction to Data Mining o Data Mining definitions o Data Mining activities o DM process o DM challenges	1	T1: Ch-1
3 4	• Data Preprocessing o Data Quality o Data preprocessing requirements o Data preprocessing techniques	2	T1: 2.1, 2.2 T2- Ch-3
5 6	• Data Exploration o Statistical descriptions of data o Measuring data similarity & dissimilarity	3	T2: Ch-2
7 8 9 10 11 12	• Classification and Prediction o Concepts of classification and prediction o Decision trees for classification o Rule based classification, o Evaluation of classification techniques o Prediction Techniques	4	T2 – 8.1, 8.2, 8.4, 8.5
13 14 15 16	• Association Analysis o Association analysis concepts o Apriori Algorithm for frequent itemsets o FP-Tree technique for frequent itemsets o Mining association rules	5	T2: Ch-6

17 18 19 20 21 22	• Clustering o Cluster analysis concepts. o Partitioning methods o Hierarchical methods for cluster analysis o Density based methods for cluster analysis o Evaluation of clustering algorithms	6	T2: 10.1, 10.2, 10.3, 10.4, 10.6
23 24	• Anomaly Detection o Concepts of Outliers o Statistical approaches o Proximity and Density based outlier detection	7	T2: 12.1,12.2,12.3, 12.4.1,12.4.3
25 26 27 28	• Data mining on unstructured (Big) data o Graph Mining methods and applications o Multimedia Data Mining o Text Mining, Web and o Social Media Mining	8	T2 (Second Edition) : 9, 10
29 30	• Data Mining Applications o Recommendation systems o Fraud Detection o Sentiment Analysis	9	T2: 13.3 R4 Ch 9 R5
31 32	• Review		

 # The above contact hours and topics can be adapted for non-specific and specific WILP programs depending on the requirements and class interests.

Select Topics for experiential learning

Topic No.	Select Topics in Syllabus for experiential learning
1	Data Preprocessing
2	Classification
3	Regression
4	Clustering

Evaluation Scheme

Legend: EC = Evaluation Component

No	Name	Type	Duration	Weight	Day, Date, Session, Time
EC-1	Assignment	Implementation based		10%	To be announced
	Quiz-I	MCQs	1 hour	5%	To be announced
	Quiz-II	MCQs	1 hour	5%	To be announced
EC-2	Mid-Semester Test	TBA	2 hours	30%	To be announced
EC-3	Comprehensive Exam	Open Book	3 hours	50%	To be announced

Note *- Evaluation components can be tailored depending on the proposed model.*

Important Information

Syllabus for Mid-Semester Test (Closed Book): Topics in Weeks 1-8
Syllabus for Comprehensive Exam (Open Book): All topics given in plan of study

Evaluation Guidelines:

1. EC-1 consists of one Assignment and two Quizzes. Announcements regarding the same will be made in a timely manner.
2. For Closed Book tests: No books or reference material of any kind will be permitted. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
3. For Open Book exams: Use of prescribed and reference text books, in original (not photocopies) is permitted. Class notes/slides as reference material in filed or bound form is permitted. However, loose sheets of paper will not be allowed. Use of calculators is permitted in all exams. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
4. If a student is unable to appear for the Regular Test/Exam due to genuine exigencies, the student should follow the procedure to apply for the Make-Up Test/Exam. The genuineness of the reason for absence in the Regular Exam shall be assessed prior to giving permission to appear for the Make-up Exam. Make-Up Test/Exam will be conducted only at selected exam centres on the dates to be announced later.

It shall be the responsibility of the individual student to be regular in maintaining the self-study schedule as given in the course handout, attend the lectures, and take all the prescribed evaluation components such as Assignment/Quiz, Mid-Semester Test and Comprehensive Exam according to the evaluation scheme provided in the handout.