FACULTY OF ENGINEERING & TECHNOLOGY

Effective from Academic Batch: 2022-23

Programme: **Master of Technology (Computer Engineering)**

Semester: **I**

Course Code: **202340105**

Course Title: **Data Mining**

Course Group: **Program Elective I**

Course Objectives: This Course is aimed to learn concepts and techniques of Data mining and Data Warehousing. The aim is to study the overall framework of data mining and understand the role of data warehousing in data analysis. The aim is also to study important techniques of data mining such as association rule mining, classification, clustering and advances in data mining with applications.

Teaching & Examination Scheme:

Contact hours per week			Course Credits	Examination Marks (Maximum / Passing)				
				Internal		External		Total
Lecture	Tutorial	Practical		Theory	J/V/P*	Theory	J/V/P*	
3	0	2	4	50 / 20	25 / 10	50 / 20	25 / 10	150 / 60

* **J**: Jury; **V**: Viva; **P**: Practical

Detailed Syllabus:

Sr.	Contents	Hours
1	**Introduction to Data Mining:** Introduction, Importance of Data Mining, Kinds of Data, Kinds of Pattern, Data Mining Technologies, Data Mining Applications, Issues in Data Mining	4
2	**Know your Data:** Data Objects and Attribute types, Statistical descriptors of data, Data visualization, Measuring data similarity and dissimilarity	4
3	**Data Preprocessing:** Overview, Data cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization	4
4	**Data Warehousing, Online Analytic Processing and Data lake**: Data Warehouse Basic Concepts, Data Warehouse Modeling: Data cube and OLAP, Data lake: Definition, The role of data lakes with big data, Data lake in comparison with data warehouse, Data lake architecture and uses of it, Data lake challenges and risks	7
5	**Frequent Pattern Mining:** Basic concepts, Market basket analysis, Apriori Algorithm, Generating Association rules from Frequent itemset, Improving efficiency of Apriori, Pattern-growth	6

	approach for Mining frequent itemsets, Pattern evaluation methods, Mining Multilevel association rules, Mining Multidimensional associations	
6	**Classification**: Basic concept, Decision tree Induction, Bayes Classification Method, Rule-based classification, Rule-based Classification, Model evaluation and selection, Techniques to improve accuracy	**6**
7	**Clustering**: Basic concepts of cluster analysis, Partitioning methods, Hierarchical Methods, Density-based methods, Evaluation of Clustering	**5**
8	**Advance Data Mining:** Applications and Trends in Data Mining, Text Mining, Web Mining, Outlier Detection	**4**

List of Practicals / Tutorials:

1	Perform data preprocessing tasks for data cleaning, smoothing and missing data.
2	Study of data visualizations using various graphs.
3	Develop an application to demonstrate OLAP operations like roll up, drill down, slice, dice.
4	Implementation of Apriori algorithm.
5	Study and working with FP-growth algorithm.
6	Working with decision tree classification technique using suitable tool.
7	Working with other classification techniques using appropriate tool.
8	Working with clustering techniques using suitable tool and techniques.
9	Study and working with text mining applications.
10	Working with various data mining tools-1
11	Working with various data mining tools-2

Reference Books:

1	J. Han, M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann
2	M. Kantardzic, "Data mining: Concepts, models, methods and algorithms, John Wiley & Sons Inc.
3	PaulrajPonnian, "Data Warehousing Fundamentals", John Willey.
4	M. Dunham, "Data Mining: Introductory and Advanced Topics", Pearson Education.
5	G. Shmueli, N.R. Patel, P.C. Bruce, "Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner", Wiley India.
6	Introduction to Data Mining – Pang-Ning Tan, Michael Steinbach and Vipin Kumar, Pearson education.
7	Arun K. Pujari, "Data Mining", University Press
8	George M.Marakas, "Mordern Data Warehousing, Data Mining and Visualization", Pearson
9	Vikram Puri and P.RadhaKrishana, "Data Mining", Oxfrod Press

Supplementary learning Material:

1	NPTEL Data Mining: https://nptel.ac.in/courses/106/105/106105174/

Pedagogy:
- · Direct classroom teaching
- · Audio Visual presentations/demonstrations

- · Assignments/Quiz
- · Continuous assessment
- · Interactive methods
- · Seminar/Poster Presentation
- · Course Projects

Internal Evaluation:

The internal evaluation comprised of written exam (40% weightage) along with combination of various components such as Certification courses, Assignments, Mini Project, Group activity, Seminar, Poster Presentation, Unit test, Quiz, Class Participation, Attendance etc. where individual component weightage should not exceed 20%.

Suggested Specification table with Marks (Theory) (Revised Bloom's Taxonomy):

Distribution of Theory Marks						**R**: Remembering; **U**: Understanding; **A**: Application, **N**: Analyze; **E**: Evaluate; **C**: Create
R	**U**	**A**	**N**	**E**	**C**	
20%	35%	20%	10%	10%	5%	

Note: This specification table shall be treated as a general guideline for students and teachers. The actual distribution of marks in the question paper may vary slightly from above table.

Course Outcomes (CO):

Sr.	Course Outcome Statements	%weightage
CO-1	Understand the functionality of the various data mining and data warehousing component.	**15**
CO-2	Appreciate the strengths and limitations of various data mining and data warehousing models.	**10**
CO-3	Demonstrate the working of algorithms for data mining tasks such as association rule mining, classification, clustering and advanced data mining techniques.	**45**
CO-4	Emphasize hands-on experience working with all real data sets.	**20**
CO-5	Compare different approaches of data warehousing and data mining with various technologies.	**10**

Curriculum Revision:	
Version:	2.0
Drafted on (Month-Year):	June-2022
Last Reviewed on (Month-Year):	-
Next Review on (Month-Year):	June-2025