FACULTY OF ENGINEERING & TECHNOLOGY

Effective from Academic Batch: 2022-23

Programme: **Bachelor of Technology (Artificial Intelligence (AI) & Data Science)**

Semester: **V**

Course Code: **202045615**

Course Title: **Data Warehousing and Data Mining**

Course Group: **Professional Elective Course**

Course Objectives: This course provides the knowledge of basic applications, concepts, and techniques of data warehousing and data mining. It introduces the concept of Data Mining as an important tool for enterprise data management and as a cutting-edge technology for building competitive advantage. The course is driven from the engineering perspective.

Teaching & Examination Scheme:

Contact hours per week			Course Credits	Examination Marks (Maximum / Passing)				
				Theory		J/V/P*		Total
Lecture	Tutorial	Practical		Internal	External	Internal	External	
3	0	2	4	50 / 18	50 / 17	25 / 09	25 / 09	150 / 53

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

Detailed Syllabus:

Sr.	Contents	Hours
1	**Overview of Data Warehousing and Business Intelligence:** What is data warehousing? Definition, 3 tier Architecture of DW Need for data warehousing, Basic concepts, Data warehouses and data marts, data warehouse metadata, Data Warehouse Modeling: Data Cube, Schema, OLTP vs. OLAP, OLAP Operations, OLAP Server Architectures, ROLAP versus MOLAP versus HOLAP, Introduction to BI, Integrating BI and DW, BI Users, Application of BI, BI Challenges	**07**
2	**Introduction to Data Mining:** Motivation for Data Mining, Definition and Functionalities, Classification of DM Systems, kind of data used for mining, Data mining models, DM task primitives, Issues in DM, KDD Process, Application of Data Mining	**05**
3	**Data Preprocessing:** Motivation behind preprocessing, data cleaning, data integration, data reduction, data transformation, data discretization and concept hierarchy generation, feature extraction, feature transformation, feature selection, introduction to Dimensionality Reduction	**06**

4	**Mining Frequent Patterns, Associations and Correlations:** Market basket analysis, Frequent Itemsets, Closed Itemsets, and Association Rules, Apriori Algorithm, Generating Association Rules from Frequent Itemsets, Improving the Efficiency of Apriori, Pattern-Growth Approach for Mining Frequent Itemsets, Pattern evaluation methods, Associative Classification	**06**
5	**Classification** General Approach to Classification, Decision Tree Induction, Bayes Classification Methods, Rule-Based Classification, Classification by Backpropagation, Support Vector Machines	**06**
6	**Clustering** Cluster Analysis, Partitioning Methods, Hierarchical Methods, Density-Based Methods, Strengths and Weakness; Outlier Detection, Clustering high dimensional data.	**05**
7	**Advance topic on Data mining:** Web Mining, Text data Mining, Spatial Data Mining, Temporal Mining, Privacy Preservation	**05**

List of Practicals / Tutorials:

1	Implement "Data Cleaning" Smoothing by binning techniques mean, median and boundaries.
2	Find the correlation for numerical data tuple using formula. Find the correlation for discrete data tuple using formula of $\chi 2$ (chi square) Analysis.
3	Implement "Data Transformation" by • Min- max normalization • Z- score normalization
4	Implement Schemas of Data warehouse.
5	Introduction to the WEKA machine learning toolkit and show data preprocessing in it.
6	Use WEKA tool to generate Association Rules using the Apriori Algorithm.
7	Explore data mining tool: DB miner
8	Explore data mining tool: Orange.
9	Study of BI tools: PowerBI
10	Study of BI tools: Tableau

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	M. Dunham, "Data Mining: Introductory and Advanced Topics", Pearson Education.
4	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
5	Ning Tan, Vipin Kumar, Michael Steinbanch Pang, "Introduction to Data Mining", Pearson Education
6	G.K. Gupta , "Introduction to Data Mining with Case Studies",PHI Learning

Supplementary learning Material:	
1	Coursera Pattern Discovery in Data Mining by Jiawei Han (https://www.coursera.org/learn/datapatterns?specialization=data-mining)
2	NPTEL - Swayam Courses: Data mining by Prof. Pabitra Mitra, IIT Kharagpur

Pedagogy:

- Direct classroom teaching
- Audio Visual presentations/demonstrations
- Assignments/Quiz
- Continuous assessment
- Interactive methods
- Seminar/Poster Presentation
- Industrial/ Field visits
- Course Projects

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

Distribution of Theory Marks in %						**R**: Remembering; **U**: Understanding; **A**: Applying; **N**: Analysing; **E**: Evaluating; **C**: Creating
R	**U**	**A**	**N**	**E**	**C**	
15%	20%	30%	20%	10%	--	

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	To demonstrate an understanding of the importance of data mining and the principles of business intelligence.	**20**
CO-2	To organize and prepare the data needed for data mining using preprocessing techniques	**30**
CO-3	To implement the appropriate data mining methods like Frequent Pattern mining on large data sets.	**30**
CO-4	To define and apply metrics to measure the performance of various data mining algorithms.	**20**

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