CVM UNIVERSITY
Aegis: Charutar Vidya Mandal (Estd.1945)

FACULTY OF ENGINEERING & TECHNOLOGY

Effective from Academic Batch: 2020-21

Programme: **Bachelor of Engineering (Information Technology)**

Semester: **VI**

Course Code: **102046706**

Course Title: **Data Mining and Business Intelligence**

Course Group: **Professional Elective Course - II**

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	40 / 14	60 / 21	20 / 7	30 / 10	150 / 52

* **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	06
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:** 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	**Concept Description, Mining Frequent Patterns, Associations and Correlations:** Concept description, Data Generalization and summarization-based characterization, Attribute relevance - class comparisons, 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:** Basic Concepts, Decision Tree Induction, Bayes Classification methods, Rule based classification, Metrics for Evaluating Classifier Performance, Cross validation, Bootstrap, Ensemble method, Bagging, boosting, Random forest	**06**
6	**Cluster Analysis:** Clustering Overview, Partitioning Clustering , K-Means Algorithm, K-Medoids, Hierarchical Clustering – Agglomerative Methods and divisive methods, Basic Agglomerative Hierarchical Clustering, Density based methods, Grid based methods, Evaluation of Clustering, Outlier Detection	**06**
7	**Advance topic on Data mining:** Web Mining, Text data Mining, Spatial Data Mining, Temporal Mining, And Multimedia Mining, information privacy and data mining	**03**
8	**Application of DM** Data mining for business applications like Balanced Scorecard, Fraud Detection, Clickstream Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance and CRM etc.	**02**
	Total	**40**

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. $$r_{A,B} = \frac{\sum(A-\bar{A})(B-\bar{B})}{(n-1)\sigma_A\sigma_B} = \frac{\sum(AB)-n\bar{A}\bar{B}}{(n-1)\sigma_A\sigma_B}$$ Find the correlation for discrete data tupel using formula of χ2 (chi square) Analysis.
3	Implement “Data Transformation” by • Min- max normalization • Z- score normalization
4	Implement Schemas of Datawarehouse.
5	Introduction to the WEKA machine learning toolkit and show data preprocessing in it.
6	Use WEKA tool and show how classification and clustering can be done.
7	Use WEKA tool to generate Association Rules using the Apriori Algorithm.
8	Explore data mining tool: DB miner
9	Explore data mining tool: Orange.
10	Case study on DM

11	Introduction to any BI tool (Qliksense, PowerBI, Tableau, etc.)
12	Mini Project

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

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**: Analyzing; **E**: Evaluating; **C**: Creating
R	**U**	**A**	**N**	**E**	**C**	
15%	25%	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 pre preprocessing techniques	**30**

CO-3	To implement the appropriate data mining methods like classification, clustering, or 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:	1.0
Drafted on (Month-Year):	June-2020
Last Reviewed on (Month-Year):	-
Next Review on (Month-Year):	June-2025