Course code	Course Title	L	T	P	C
BCSE208L	Data Mining	2	0	0	2
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To introduce the fundamental processes data warehousing and major issues in data mining.
2. To impart the knowledge on various data mining concepts and techniques that can be applied to text mining, web mining etc.
3. To develop the knowledge for application of data mining and social impacts of data mining.

Course Outcome

Upon completion of the course the student will be able to

1. Interpret the contribution of data warehousing and data mining to the decision-support systems.
2. Construct the data needed for data mining using preprocessing techniques.
3. Discover interesting patterns from large amounts of data using Association Rule Mining.
4. Extract useful information from the labeled data using various classifiers and Compile unlabeled data into clusters applying various clustering algorithms.
5. Demonstrate capacity to perform a self-directed piece of practical work that requires the application of data mining techniques.

Module:1	Data Warehousing	4 hours

Introduction to Data warehouse - Data Warehouse models- Data warehouse architecture: Three-tier data warehouse architecture - Data warehouse modeling: Data cube and OLAP – Star and Snowflake Schema.

Module:2	Introduction to Data Mining	3 hours

Introduction to data mining - Data mining functionalities - Steps in data mining process- Classification of data mining systems - Major issues in data mining.

Module:3	Data Preprocessing	3 hours

Data Preprocessing: An overview - Data cleaning - Data integration -Data reduction - Data transformation.

Module:4	Frequent Pattern Mining	4 hours

Frequent Pattern Mining: Basic Concepts and a Road Map - Efficient and scalable frequent item set mining methods: Apriori algorithm, FP-Growth algorithm - Mining frequent item sets using vertical data format.

Module:5	Classification Techniques	5 hours

General approach to classification - Classification by decision tree induction - Bayes classification methods - Model evaluation and selection - Techniques to improve classification accuracy - advanced classification methods: Bayesian belief networks- Lazy learners.

Module:6	Cluster Analysis	5 hours

Types of data in cluster analysis - Partitioning methods - K Medoid Clustering - Density based methods - Grid based methods - Outlier analysis.

Module:7	Data Mining Trends and Research Frontiers	4 hours

Overview of Web mining-Temporal and Spatial mining-Other methodologies of data mining: Statistical data mining- Data mining applications.

Module:8	Contemporary Issues	2 hours

	Total Lecture hours:	**30 hours**
Text Book(s)		
1.	Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, third edition, 2013.	
Reference Books		
1.	Parteek Bhatia, Data Mining and Data Warehousing: Principles and Practical Techniques, Cambridge University Press, 2019.	
2.	Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar, Introduction to Data Mining, Pearson, 2nd Edition, 2019.	
Mode of Evaluation : Continuous Assessment Tests, Quizzes, Assignment, Final Assessment Test		
Recommended by Board of Studies	12-05-2022	
Approved by Academic Council	No. 66	Date 16-06-2022