Course Code CSA4003	Data Mining And Data Warehousing	Course Type	LT
		Credits	3

Course Objectives:

- Understand the basic concepts of data mining.
- Apply the data mining functionalities
- Assess the strengths and weaknesses of various data mining techniques.

Course Outcomes:

At the end of the course, students should able to

- Implement the data warehouse architecture
- Explain the functionalities of data mining
- Explore the different data preprocessing techniques
- Identify the association rules using frequent itemset mining algorithms
- Describe the classification and clustering techniques.

Student Outcomes (SO): a, b, c, l

a. An ability to apply the knowledge of mathematics, science and computing appropriate to the discipline.

b. An ability to analyze a problem, identify and define the computing requirements appropriate to its solution.

c. An ability to design, implement and evaluate a system / computer-based system, process, component or program to meet desired needs.

l. An ability to apply mathematical foundations, algorithmic principles and computer science theory in the modelling and design of computer-based systems (CS).

Module No.	Module Description	No.of Hours	SO
1	**Data Warehouse:** Basic Concepts, Differences between Operational Database Systems and Data Warehouses- A Multitiered Architecture - Data Warehouse Models : Extraction, Transformation and Loading - Metadata Repository -Data Cube and OLAP -Data Warehouse Design and Usage – Data warehouse implementation.	7	a, c
2	**Introduction To Data Mining:** Introduction - The evolution of database system technology - Steps in knowledge discovery from database process - Architecture of a data mining systems - Data mining on different kinds of data - Different kinds of pattern - Technologies used - Applications - Major issues in data mining - Classification of data mining systems - Data mining task primitives - Integration of a data mining system with a database or data warehouse system.	9	a, b
3	**Data Preprocessing:** Data Objects and attribute types - Basic statistical description of data - Data visualization – Measuring data similarity and dissimilarity - Data cleaning - Integration - Data reduction - Data transformation and data discretization.	10	c, l
4	**Association Rule Mining:** Basic concepts - Frequent itemset mining methods - Apriori algorithm, APattern growth approach for mining frequent itemsets,Mining frequent itemsets using vertical data format, Mining closed and max patterns - Pattern mining in multilevel and	9	b,c

	multidimensional space – Constraint based Frequent pattern mining - Mining High-Dimensional Data and Colossal Patterns.		
5	**Classification And Clustering**: Classification : Basic concepts - Decision tree induction - Bayes classification methods-Rule BasedClassification- Model Evaluation and Selection - Techniques to Improve Classification Accuracy -Bayesian Belief Networks - Classification by Backpropagation - Cluster Analysis – Partitioning methods- Hierarchical methods.	10	b, c
5	**Guest Lecture on Contemporary Topics**	**2**	
	Total Hours:	**45**	

Mode of Teaching and Learning: *Flipped Class Room, Activity Based Teaching/Learning, Digital/Computer based models, wherever possible to augment lecture for practice/tutorial and minimum 2 hours lectures by industry experts on contemporary topics*

Mode of Evaluation and assessment:
The assessment and evaluation components may consist of unannounced open book examinations, quizzes, student's portfolio generation and assessment, and any other innovative assessment practices followed by faculty, in addition to the Continuous Assessment Tests and Term End Examinations.

Text Book(s):

1.	Jiawei Han, Micheline Kamber and Jian Pai, Data Mining: Concepts and Techniques, Morgan Kauffman, 2013.
2.	Alex Berson and Stephen J Smith, Data Warehousing, Data Mining, and OLAP, Mcgraw- Hill,2008.

Reference Book(s):

1.	David Hand, Heikki Manila, PadhraicSymth, Principles of Data Mining, MIT Press, 2004
2.	Margaret H.Dunham, Data Mining: Introductory and Advanced Topics, Pearson Education 2008

Recommendation by the Board of Studies on	**17.01.2020**
Approval by Academic council on	**20.01.2020**
Compiled by	**Dr S Sountharrajan**