Detailed Syllabus
Lecture-wise Breakup

Course Code	**15B22CI621**	**Semester** : **Even**	**Semester** 6th **Session** 2021 -2022 **Month from Jan 22 to June 22**
Course Name	**Data Mining And Web Algorithms**		
Credits	4	**Contact Hours**	4(3+1)

Faculty (Names)	**Coordinator(s)**	Archana Purwar
	Teacher(s) (Alphabetically)	Archana Purwar

COURSE OUTCOMES		**COGNITIVE LEVELS**
C313.1	Understand the basics of data mining and pre-processing of data.	Understand Level (Level 2)
C313.2	Analyze the transactional data for finding frequent and interesting patterns using association rule mining techniques like Apriori and FP-Growth.	Analyse Level (Level 4)
C313.3	Apply a wide range of classification techniques like Naïve-bayes, decision tree, and KNN for the numerous application including fraud detection, target marketing, medical diagnosis, etc.	Apply Level (Level 3)
C313.4	Cluster the similar/dissimilar objects using different methods like partitioning, hierarchical and density based clustering.	Create Level (Level 6)
C313.5	Analyze the link structure of web using page rank and HITS algorithms.	Analyse Level (Level 4)
C313.6	Develop recommendation system using collaborative filtering techniques	Create Level (Level 6)

Module No.	**Title of the Module**	**Topics in the Module**	**No. of Lectures for the module**
1.	**Course overview**	What Motivated Data Mining? Why Is It Important? What Is Data Mining? Data Mining—On What Kind of Data? Data Mining Functionalities—What Kinds of Patterns Can Be Mined? Are All of the Patterns Interesting? Data mining process, Types of datasets and attributes, Major Issues in Data Mining.	**03**
2.	Data Preprocessing	Getting To know your data, Data extraction, Data cleaning, Data Integration and transformation, Data reduction	**06**
3.	Association Rule mining	Usability and Complexity Analysis of Apriori Algorithm, Sampling Algorithm, Partitioning, Using multiple minimum supports	**05**
4.	Classification Algorithms	Issues Regarding Classification and Prediction, Bayesian Classification, Usability and Complexity Analysis of Bayesian algorithm, Nearest Neighbor algorithm, Decision Tree based algorithm.	**07**
5.	Clustering Algorithms	Clustering Algorithms: Types of Data in Cluster Analysis, Similarity Measures, A Categorization of Major Clustering Methods, Partitioning Methods, Hierarchical Methods, Usability and Complexity Analysis of Agglomerative	**10**

		Hierarchical Algorithm, K-means and K-Mediod Partitioning Algorithm, subspace clustering techniques, ,Applications of clustering.	
6.	Web algorithms:	Web algorithms: Link Based Search Algorithm, Web Crawling, Indexing, Searching, Zone Indexing, Term-Frequency, Link Analysis Algorithm.	**03**
7.	Ranking Algorithms:	Ranking Algorithms: Page rank, Hits ranking algorithms	**03**
8	Web caching Algorithm :	Web caching Algorithm : LRV, FIFO, LRU, Random, OPT	**02**
9	Recommendation Algorithms:	Recommendation Algorithms: Collaborative Filtering, Item-to-Item recommendation, Memory Based Recommendation,	**03**
		Total number of Lectures	**42**

Evaluation Criteria

Components	**Maximum Marks**
T1	20
T2	20
End Semester Examination	35
TA	25 [Attendance /Assignments/Mini-project]
Total	**100**

Project based learning: Each student in a group of 3-4 will have to develop a mini project based on association mining approaches, classification methods, page rank as well as HITS algorithm and recommendation algorithm. The students can choose any real-world application that requires some decision-making. The students have to implement the mini-project using any open-source programming language. Project development will enhance the knowledge and employability of the students in IT sector.

Recommended Reading material: Author(s), Title, Edition, Publisher, Year of Publication etc. (Text books, Reference Books, Journals, Reports, Websites etc. in the IEEE format)

1.	**Text Books**
2.	Han, Jiawei, Jian Pei, and Micheline Kamber. Data mining: concepts and techniques. Elsevier, 3rd edition ,2012
3.	Kimball R. and Ross M ,The Data Warehouse Toolkit", Wiley, 3rd edition,2013
4.	Pujari, Arun K, Data mining techniques , Universities press, 3rd edition , 2013
5.	Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, second edition, 2019
6.	Langville, Amy N., and Carl D. Meyer. Google's PageRank and beyond: The science of search engine rankings. Princeton University Press, 2012.
	Reference Books
7.	Soumen Chakrabarti, Mining the Web: Discovering knowledge from hypertext data", Morgan Kaufmann, Elsevier,2002
8.	Berson, Alex, and Stephen J. Smith. Data warehousing, data mining, and OLAP. McGraw-Hill, Inc., 2004
9.	Inmon W.H.,Building the Data Warehouse ,4th Edition, Wiley,2005
10.	Anahory, Sam, and Dennis Murray. Data warehousing in the real world: a practical guide for building

	decision support systems. Addison-Wesley Longman Publishing Co., Inc., 1997.
11.	Dunham, Margaret H. Data mining: Introductory and advanced topics. Pearson Education India, 2006.
12.	Mattison, Rob, and Brigitte Kilger-Mattison. Web warehousing and knowledge management. McGraw-Hill School Education Group, 1999.
13.	Hand, David, Heikki Mannila, and Padhraic Smyth. Principles of data mining.PHI, 2005
14.	C.D. Manning, P. Raghavan, H. Schütze., Introduction to Information Retrieval, Cambridge Press,1st edition, 2008.