COURSE CODE: **CSC3066**	L-T-P: **4-1-1**
COURSE NAME: **DATA MINING AND WAREHOUSING**	CONTACT HOURS/WEEK: **7**
	TOTAL MARKS: **100 (**INTERNAL: **60,**
COURSE TYPE: **ELECTIVE/OPEN**	EXTERNAL: **40)**
NUMBER OF CREDITS: **6**	NATURE: **GRADED**

COURSE OBJECTIVES:
1. To introduce students the basic concepts of Data Warehouse and techniques and applications of Data Mining.
2. To develop skills for designing and implementing systems for data mining to solve practical problems in a variety of disciplines.
3. To provide students the experience of doing independent study and research.

COURSE PREREQUISITE:
- Programming knowledge of C, C++.
- Basic knowledge of Mathematics-Statistics.
- Basic concepts of Database.

COURSE OUTCOMES:

At the end of the course, students will be able to:
- explain the components and architecture of data warehouse architecture
- Illustrate different data mining techniques such as association rule mining, clustering and classification.
- Analyze different data mining algorithms such as K-means, DBSCAN, FR-tree growth, A priori, CURE, BIRC, ROCK, CART, C4.5 etc.
- Analyze the uses of developing areas-web mining, text mining and sequential data mining.

COURSE CONTENT:

Unit No & Name	Components of the Unit	No of contact hours	Marks
UNIT-I: **Introduction to Data Mining**	- **Basic Concepts:** Data Mining, kinds of patterns that can be mined, Data Mining versus Database systems, Data preparation, cleaning and visualization. - **Data Warehousing:** Differences between database systems and Data Warehouse, Data Warehouse architecture and its components, Warehouse versus Data Mining (OLTP & OLAP), OLAP tools,	20	30

	Data cubes, Multidimensional Data.		
UNIT-II: **Data Mining Techniques**	• **Association Rules:** What is an association rule? Mining association rules, frequent sets and border sets, algorithms for mining association rules – A priori algorithm, Pincer-search algorithm, Border algorithm, FP-tree growth algorithm, generalized association rule, association rule with item constraints. • **Clustering:** Hierarchical versus Partitional clustering, types of data in clustering, Partitional algorithms – k-means, k-mediods, PAM, CLARA, CLARANS. Density based clustering algorithm – DBSCAN. Hierarchical algorithms – BIRCH, CURE. Categorical clustering algorithms – ROCK, CACTUS. • **Decision Trees :** Introduction, tree construction principle, decision tree generation algorithms – CART, ID3, C4.5 • **Other techniques for Data Mining :** Concepts of Genetic algorithms, Artificial Neural Network and Rough sets and their application in the domain of data mining. Introduction to Web Mining, Text Mining, Temporal data mining.	70	70
	TOTAL	90	100

TEXTBOOKS/ RECOMMENDED READINGS:

- Puzari K.; *Data Mining Techniques*; University Press

- Han J., Kamber M.; *Data Mining Concepts and Techniques*; India Morgan Kaufmann & Harcourt

- Soman K. P., Diwakar S., Ajay V.; (2008); *Insight into Data Mining*: *Theory and Practice* ; P.H.I (Eastern Economy Edition

- Jain K. and Dukes R. C.; *Algorithms for Clustering Data*; Prentice-Hall

- Cios K., Pedrycz W., Swiniarski R; (1998); *Data Mining : Methods of Knowledge Discovery*; Boston Kluwer Academic Publishers, ,

COURSE ASSESSMENT DETAILS:

Internal assessment: Class tests, Assignments, Laboratory tests, Seminar
External assessment: End Semester Examination

LESSON PLAN

Subject Name	:	**Data Mining and Data Warehousing**		
Paper Code	:	**CSC3066/INF3066**	Session:	**2019-2020**
Program Name	:	**M.Sc. (CS/IT)**	Semester:	**Third**
Faculty Name	:	**Dwipen Laskar**		
Date	:	**July, 2019 to December, 2019**		

Detailed Lesson Plan

UNIT-I (Introduction to Data Mining)

Lecture No	Topics to be Covered
1	**Basic Concepts:** Data Mining, Definition, kinds of patterns that can be mined,
2	KDD vs Data Mining, DBMS vs DM, Other Related areas
3	Data Mining versus Database systems, Data preparation
4	Issues and Challenges in DM, Application areas of DM
5	Data cleaning and visualization.
6	**Data Warehousing:** Differences between database systems and Data Warehouse
7	Data Warehouse architecture and its components
8	Warehouse Server, Meta Data, Warehouse versus Data Mining
9	OLTP Engines & OLAP
10	OLAP tools
11	Data cubes
12	Multidimensional Data

UNIT-II (: Data Mining Techniques)

13	**Association Rules:** What is an association rule? Mining association rules, Methods to discover association rules
11	Frequent sets and border sets
12	Algorithms for mining association rules – A priori algorithm

13	Partitioned Algorithm
14	Pincer-search algorithm
15	Dynamic Item set Counting Algorithm
16	Border algorithm
17	FP-tree growth algorithm
18	Rapid Association Rule Mining, Eclat and dEclat
19	Generalized association rule
20	Association rule with item constraints
21	**Clustering:** Clustering Paradigms, Hierarchical versus Partitioned clustering
22	Types of data in clustering
23	Partitional algorithms – k-means
24	Partitional algorithms – k-mediods
25	Partitional algorithms – PAM
26	Partitional algorithms – CLARA
27	Partitional algorithms – CLARANS
28	Density based clustering algorithm – DBSCAN
29	Hierarchical algorithms – BIRCH
30	Hierarchical algorithms – CURE
31	Categorical clustering algorithms – ROCK
32	Categorical clustering algorithms –CACTUS
33	clustering algorithms –STIRR
34	**Decision Trees :** Introduction, tree construction principle, Best Split
35	Splitting Indices, Splitting Criteria
36	DECISION tree generation algorithms – CART
37	DECISION tree generation algorithms –ID3
38	DECISION tree generation algorithms –C4.5
39	DECISION tree generation algorithms –CHAID, RainForest
40	DECISION tree generation algorithms –CLOUDS, BOAT

41	Pruning Techniques, Integration of Pruning and Construction
42	**Other techniques for Data Mining** : Concepts of Genetic algorithms, Foundation of Genetic Algorithms, Search space
42	Operators of Genetic Algorithms: Selection, Crossover, Mutation, Application of Genetic Algorithms
43	Artificial Neural Network: Perceptron, Characteristics of ANN , Multilayer ANN
44	Linear Support Vector Machine: Separable and non Separable
45	Non-linear SVM, characteristics of SVM
46	Rough sets: concepts and Definition
47	REDUCT in Rough Set, Rule Extraction
48	Rough Set and Fuzzy Sets, Application of Rough Sets in the domain of data mining
49	Introduction to Web Mining , Web Content Mining, Page Rank, Web Usage Mining
50	Text Mining, Unstructured Test, Episode Rule Discovery for Texts, Text Clustering
51	Temporal data mining, Temporal Association Rules
52	Sequence Mining, GSP Algorithm
53	SPADE (Sequential Patter Discovery using Equivalence Classes)
54	SPIRIT (Sequential Pattern Mining with Regular expression Constarints), WUM (Web Utilization Miner)

(**Dwipen Laskar**)
(Assistant Professor, Dept. of Computer Sc., GU)