LESSON PLANS

FOR

ODD SEMESTER

(July, 2019 to December, 2019)

EVEN SEMESTER

(Jan, 2020 to June, 2020)

SESSIONS

2019-2020

COURSE CODE**: CSC3066**	L-T-P**: 4-1-1**
COURSE NAME**: DATA MINING AND WAREHOUSING**	CONTACT HOURS/WEEK**: 7**
COURSE TYPE**: ELECTIVE/OPEN**	TOTAL MARKS**: 100 (**INTERNAL**: 60,** 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

DEPARTMENT OF COMPUTER SCIENCE

Gopinath Bordoloi Nagar, Gauhati University

Guwahati-781014, Assam, India

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)

COURSE CODE**: CSC2036**	L-T-P**: 4-1-1**
COURSE NAME**: SOFTWARE ENGINEERING**	CONTACT HOURS/WEEK**: 7**
COURSE TYPE**: CORE**	TOTAL MARKS**: 100 (**INTERNAL**: 60,** EXTERNAL**: 40)**
NUMBER OF CREDITS**: 6**	NATURE**: GRADED**

COURSE OBJECTIVES:

1. To provide students the knowledge of SE challenges, Software process, S/W development process model and problem analysis.
2. The give students the concepts of role of software architecture, architecture views, software cost estimation model, quality plan, and risk management.
3. To familiarize students with concepts of module level concept, OO Analysis and OO Design, UML, Coding process, refactoring, verification, testing fundamentals.

COURSE PREREQUISITE:

- Basic knowledge of database management system.

COURSE OUTCOMES:

At the end of the course, students will be able to:

- Compare different software development processes and their challenges.
- Create software require specification and translate it into an implementable design, following a structured and organize process.
- Implement different software estimation metrics such as cost, effort size, staffing etc.
- Make effective use of UML, along with design strategies such as defining software architecture, separation of concerns and design patterns.

COURSE CONTENT:

Unit No & Name	Components of the Unit	No of contact hours	Marks
UNIT-I: **Software challenges and Software process**	• Problem domain, SE challenges • SE approach. Software process, Characteristics of SW process, SW development process model.	10	10
UNIT-II: **Software requirement and specification**	• SW requirement, problem analysis, requirement specification. • Functional specification, validation, matrices.	10	15

UNITIII:**Software architecture views and cost estimation model.**	• Role of SW architecture, architecture view, component and connector view, style for C&C view. • Process planning, Effort estimation, Software Cost Estimation based on COCOMO II cost model. • Scheduling and staffing. • SW configuration management plan, quality plan, risk management, project monitoring plan.	30	35
UNIT-IV: **Design principles and Methodology**	• Design principle • Module level concept • Design notation and specification, • Structured design methodology verification. • OO Analysis and OO Design. OO Design concept, UML. OO Design methodology.	25	25
UNIT-V: **Detail Design and Testing fundamentals**	• Detail design and PDL • Verification, Metrices, Programming principles and guidelines, coding process, refactoring, verification. Testing fundamentals.	15	15
	Total	90	100

TEXTBOOKS/ RECOMMENDED READINGS:

- Jalote P. ; *An integrated Approach to Software Engineering*; Narosa Publishing House
- Patton R.; *Software Engineering;* Pearson Education.
- Agarwal K. K., Singh Y,; *Software Engineering*; New Age International Publisher.
- Sommerville I.; *Software Engineering*; Pearson Education (Addison Wesley)
- Pressman R.S.; *Software Engineering: A practitioner's Approach*; McGraw Hill.

COURSE ASSESSMENT DETAILS:

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

DEPARTMENT OF COMPUTER SCIENCE

Gopinath Bordoloi Nagar, Gauhati University

Guwahati-781014, Assam, India

LESSON PLAN

Subject Name : **Software Engineering**

Paper Code : **CSC2036/INF2036** **Session:** **2019-2020**

Program Name : **M.Sc. (CS/IT)** **Semester:** **Second**

Faculty Name : **Dwipen Laskar**

Date : **Jan, 2020 to June, 2020**

Detailed Lesson Plan

UNIT-I (Software challenges and Software process)

Lecture No	Topics to be Covered
1	Problem domain: Software vs Program, Process vs Product, Exploratory Style of Programming
2	SE Challenges, Software Crisis, Software Process Model, SE approaches
3	Characteristics of SW process, Emergence of Software Engineering, Notable Changes in Software Development Practices
4	SW development process model: basic concepts, Build and Fix model, Waterfall model, Classical Waterfall Model
5	Iterative Waterfall Model, V-Model, Prototyping Model
6	Incremental Developmental Model, Evolutionary Model
7	Spiral Model, Agile Software Development, Comparison of Different Life Cycle Models

UNIT-II (Software requirement and specification)

13	SW requirement: Requirement Gathering and Analysis, Software Requirement Specification(SRS):-users of SRS, Characteristic of Good and bad SRS
14	Functional Specification: how to identify and Document, Functional Requirements, Traceability
15	Representation of complex logic in SRS, Decision Tree and Decision Table
16	Formal Technique
17	Axiomatic Specification, validation, matrices

UNIT-III (Software architecture views and cost estimation model)

18	Role of SW architecture, architecture view, component and connector view, style for C&C view.

19	Software Project Management Complexities, Responsibilities of SW Project Manager
20	Process planning, Software Planning, Sliding Window Planning
21	Software Size Estimation: LOC, Function Point Metric, Feature Point Metric
22	Empirical Estimation Techniques, Expert Judgment, Delphi Cost Estimations
23	Effort estimation,
24	Cost Estimation based on COCOMO: Basic, Intermediate and Complete COCOMO
25	Staffing Level Estimation: Norden's Work, Putnam's Work, Jensen's Model
26	Software Scheduling and staffing: Work Breakdown Structure, CPM,
27	Activity Network, PERT Charts, Gantt Chart
28	Organization Structure: Chief Programmer team, Democratic Team and Mix Team
29	Team Structure: Project Format, Functional Format, Matrix Format
30	Risk Management, Risk Identification, Types of Risks, Risk Abatement Policies
31	SW configuration management plan, Source Code Control System, RCS, Project monitoring plan.
32	Software Reliability, Hardware vs Software Reliability, Reliability Metrics
33	Reliability Growth Model, Statistical Testing, Quality plan: McCall's Quality factors, ISO9000
34	SEI Capability Maturity Model, Comparison of SEI/CMM and ISO 9000

UNIT-IV (Design principles and Methodology)

35	Design principle, Outcome of design process, classification of design activities and design methodologies
36	Characteristics of good design, Module level concept, Design notation and specification
37	Layered architecture of modules, Cohesion and Coupling
38	Overview of SA/SD, Structured Analysis (DFD)
39	Structured Analysis (DFD): Context Level Diagram, Types of DFDs, Notations of DFD, Data dictionary
40	Transformation of DFD into Structure Chart, Detailed Design
41	OO Analysis and OO Design: OO Design concept, Origin of UML, Evolution of UML, UML Diagrams
42	UML Diagrams: Use Case Diagram, Representation, Factoring, Use Case Packaging
43	UML Diagrams: Class Diagram, Object Diagram
44	UML Diagrams: Interactive Diagrams, State Chart Diagram, Deployment and Component Diagram

45	Programming principles and guidelines, coding process, Code Walkthrough, Code Inspection and Clean Room Testing, Software Documentation
46	Software Testing: basic concepts, Unit Testing, Integration and System Testing, Acceptance Testing
47	Black Box Testing-Equivalence Partitioning, Boundary Value Analysis
48	White Box Testing-Statement Coverage, Branch Coverage
49	Condition Coverage, Path Coverage, McCabe's Cyclomatic Complexity, Data Flow based Testing, Mutation Testing
50	Debugging Approaches, Guidelines, Program Analysis Tools: Static and Dynamic

UNIT-V (Detail Design and Testing fundamentals)

52	Detail design and Process Design Language (PDL)
53	Software Verification and Validation,
54	Code Refactoring concepts and Techniques
55	Integration Testing and System Testing: Smoke Testing, Performance Testing, Error Seeding
56	Software Maintenance: Reverse Engineering, Estimation of Maintenance Cost, CASE tools

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