Data Mining & Data warehousing

BEG 476CO

Year: IV **Semester:II**

Teaching Schedule Hours/Week			Examination Scheme				
Theory	Tutorial	Practical	Internal		Final		Total
			Theory	Practical	Theory	Practical	
3	1	2	20	-	80	-	100

Goals: This course introduces advanced aspects of data warehousing and data mining, encompassing the principles, research results and commercial application of the current technologies. To introduce students to the basic concepts and techniques of Data Mining. To develop skills of using recent data mining software for solving practical problems. To gain experience of doing independent study and research.

Course Content:

Unit 1. Introduction to Data Mining **4 Hrs.**

Basic concepts of data mining
Use and benefits of data mining
Application of data mining
KDD Environment: Data selection cleaning, enrichment, coding and mining
Problems in data mining

Unit 2. Introduction to Data Warehousing **4 Hrs.**

Basic concepts of data warehousing
Use and benefits of data warehousing
Application of data warehousing
Problems in data warehousing

Unit 3. Data warehouse logical and Physical design **6 Hrs.**

Data warehouse logical design: star schemas, fact tables, dimensions, other schemas, multidimensional data models, materialized views
Data warehouse physical design: hardware and I/O considerations, parallelism, indexes

Unit 4. Data warehousing technologies and implementations **4 Hrs.**

Data extraction, transportation, transformation, loading and refreshing.

Unit 5. Data Warehouse to Data Mining **9 Hrs.**

Data mining architecture
Data warehouse architecture
OLAP architecture
Types of OLAP servers
OLAP operations in Multidimensional data models
OLAP to OLAM
Stages of Data Mining Process

Unit 6. Data Mining Approaches and Methods **10 Hrs.**

Models of Data Mining
Data Mining Techniques
Data Mining Tasks

Classification and Predictions

- Decision tree, rule-based classification, Backpropagation, genetic algorithm, Linear regression, non-linear regression

Association rules and Mining frequent patterns

- Market basket analysis, APriori algorithm, FP growth

Clustering

- Partitioning method (K Means, K Medoids)
- Hierarchical method (Agglomerative, Divisive)

Unit 7. Mining complex types of data **3 Hrs.**

Multimedia Data mining
Text mining
Web mining

- Web content mining, web usage mining, web structure mininig

Unit 8. Application and trends in data warehousing and data mining **5 Hrs.**

Integration of data mining tools with database systems
Data mining in distributed heterogeneous database systems
Importance of data mining in Marketing, E- commerce and CRM
Aspects of Security and Privacy in Data Mining
Social impact of data mining
Trends in data mining

Reference Books:

"**Data Mining Concepts and Techniques"**, Morgan Kaufmann J. Han, M Kamber, Second Edition

Sam Anahory, Dennis Murray, **"Data warehousing In the Real World",** Pearson Education.

Adriaans, P. and D. Zatinge, **" Data Mining" ,** Addison Wesley, 1996

Kimball, R., **"The Data Warehouse Toolkit",** Wiley, 1996.

W.H.Inmon, **"Building the Data Warehouse"**, 3rd Edition, Wiley, 2003.

Margaret H.Dunham, **"Data Mining: Introductory and Advanced Topics",** Pearson Education 2004.

Prerequisite: C, Data Structure, Database Management Systems