COURSE TITLE	**DATA MANAGEMENT AND DATA WAREHOUSING**
COURSE CODE	**01CB0604**
COURSE CREDITS	**3**

Objective:

1 Understand the fundamental concept and significance of data mining.
2 Recognize the key tasks involved in the data mining process.
3 Explain the architecture and components of a typical data mining system

Course Outcomes: After completion of this course, student will be able to:

1 Apply knowledge of data mining system and components to design efficient analytical workflows.
2 Apply appropriate data mining techniques based on specific data characteristics and analysis goals.
3 Analyze and implement data warehouse solutions in designing, preprocessing, and utilizing datasets.
4 Analyze data mining tools such as Entrez, BLAST, and sequence retrieval systems, and evaluate the structure and utility of biological databases.
5 Evaluate the effectiveness and suitability of biological data mining tools and databases for solving domain-specific problems.

Pre-requisite of course: Basic knowledge of Bioinformatics

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
2	0	2	50	30	20	25	25

Contents : Unit	Topics	Contact Hours
1	**Data Mining** Definition, Data mining task, Data mining process, architecture and component of a typical data mining system, Mining frequent pattern, associations, and correlation: Pattern mining and Interestingness of pattern, Association Rules	5
2	**Clustering** Portioning method, Hierarchical method, Density-Based Methods, Grid-Based Methods, Model-Based Clustering Methods, Clustering High-Dimensional Data, Constraint-Based Cluster Analysis, Outlier Analysis. Mining Stream, Time-Series, and Sequence Data	5

DR. SEJAL SAHILKRISHNA SHAH
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Contents : Unit	Topics	Contact Hours
3	**Data Warehousing** Basic Concepts, Data Warehouse Architecture, Benefits of a data warehouse, Three-tier Decision Support Systems (DSS), DataMart, Online Analytical Processing (OLAP) Engine, OLAP Servers (ROLAP, MOLAP, HOPAP), Multidimensional Data Model, Data Cube, Warehouse schema (Star schema, Snowflake schema); Enterprise Warehouse, Virtual Data Warehouse; Metadata; Data Preprocessing, Data Warehouse Design and Usage	6
4	**Application of Data Mining in Biodata analysis** DNA/protein sequence Analysis, Genome analysis, Protein Structure Analysis, Pathway analysis, microarray data analysis, annotation, gene ontology, gene mapping Introduction to biological database: Designing of biological databases, Types of biological database: Primary database, Secondary database, Composite database	5
5	**Relational database management system (RDBMS)** sequence query language (MySQL)- Overview, Tables, Queries, creating and using database Design, implementation, and updating of bioinformatics knowledge bases	7
Total Hours		**28**

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**Module1:** To perform data mining from databases of specific disease	2
2	**Module1:** To perform constraint-based association mining	2
3	**Module2:** To apply clustering methods on high-dimensional data	2
4	**Module2:** To analyze time-series data	2
5	**Module3:** To perform text and spatial data analysis	2
6	**Module3:** To apply warehouse schema Snowflake schema	2
7	**Module3:** To apply warehouse schema Star schema	2
8	**Module4:** To analyze primary, secondary, and composite biological databases	2
9	**Module4:** To analyze dimensionality reduction using PCA tools	2
10	**Module4:** To analyze dimensionality reduction using offline tools	2

DR. SEJAL SAHILKRISHNA SHAH
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
11	**Module4:** To analyse the different type of schema using online tools	2
12	**Module5:** To generate a simple three-tier Decision Support System (DSS)	2
13	**Module5:** To create a simple database using SQL basics	2
14	**Module5:** To create a database with search function feature	2
Total Hours		**28**

Textbook :

1 Data Mining Techniques, A. K. Pujari, University Press, Hyderabad, 2006

2 Data Mining: Concepts and Techniques, Jiawei Han and Micheline Kamber, Elsevier, 2000

References:

1 Bioinformatics: Sequence and Genome Analysis, Bioinformatics: Sequence and Genome Analysis, Mount, D. W, Cold Spring Harbor, 2001

2 Data mining in bioinformatics, Data mining in bioinformatics, Wang et al, Springer-Verlag, 2005

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery					
Remember / Knowledge	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Higher order Thinking / Creative**
20.00	30.00	20.00	10.00	10.00	10.00

Instructional Method:

1 The course delivery method will depend upon the requirement of content and need of students. The teacher in addition to conventional teaching method by black board, may also use any of tools such as demonstration, role play, Quiz, brainstorming, MOOCs etc.

2 The internal evaluation will be done on the basis of continuous evaluation of students in the laboratory and class-room.

3 Practical examination will be conducted at the end of semester for evaluation of performance of students in laboratory.

Supplementary Resources:

1 https://www.cdata.com/sync/

DR. SEJAL SAHILKRISHNA SHAH
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Supplementary Resources:

2 https://www.datacamp.com/courses/introduction-to-data-warehousing