

DATA WAREHOUSING& DATA MINING

(Professional Elective–I)

COURSE CODE: IT314 B

L T P E O

3 0 0 1 2

CREDITS 3

Sessional Marks: 40

End Exam Marks: 60

End Exam: 3Hours

Prerequisite(s): DBMS

Course Objectives

1. To introduce the basic concepts of Data Warehouse and Data Mining techniques.
2. Examine the types of the data to be mined and apply pre-processing methods on raw data.
3. Discover interesting patterns, analyse supervised and unsupervised models and estimate the accuracy of the algorithms.

Course Outcomes

After completion of this course, the students will be able to:

CO-1: Apply critical operations involved in modeling and designing data warehouses.

CO-2: Evaluate at a quality and implement well-known data mining techniques for obtaining interesting knowledge from data.

CO-3: Reflect on advantages and disadvantages of data mining solutions to solve real life problems.

CO-4: Evaluate the models, their usefulness and usability towards research and innovation.

Mapping of Course Outcomes with POs and PSOs

COs/POs-PSOs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	2	2	3	1	3	1	1	1	2	1		1	3	2
CO2	2	3	3	3	3	2	1	1	2	1		1	3	2
CO3	1	3	3	3	3	2	1	1	2	1		1	3	2
CO4	3	3	3	3	3	2	2	1	2	1		2	3	2

UNIT-I Data warehouse

9 Lectures

Introduction to Data warehouse, Difference between operational database systems and data warehouses, Data warehouse Characteristics, Data warehouse Architecture and its Components, Extraction-Transformation-Loading, Logical (Multi-Dimensional), Data Modeling, Schema Design, Star and Snow-Flake Schema, Fact Constellation, Fact Table, Fully Addictive, Semi-Addictive, Non-Addictive Measures; Fact-Less-Facts, Dimension Table Characteristics; OLAP Cube, OLAP Operations, OLAP Server Architecture-ROLAP, MOLAP and HOLAP.

Learning outcomes: At the end of this unit, the students will be able to

- Perform ETL Operations.
- Design the data warehouse.
- Apply OLAP Operations

UNIT-II Introduction to Data Mining

9Lectures

Fundamentals of data mining: Data Mining Functionalities, Classification of Data Mining systems, Data Mining Task Primitives, Integration of a Data Mining System with a Database or Data Warehouse System, Major issues in Data Mining.

Data Preprocessing: Need for Preprocessing the Data, Data Cleaning, Data Integration & Transformation, Data Reduction, Discretization and Concept Hierarchy Generation. Applications and Trends in Data mining

Learning outcomes: At the end of this unit, the students will be able to

- Classify data mining systems and identify the major issues in data mining.
- Apply pre-processing steps.
- Characterize real world applications of data mining.

UNIT-III Association Rules

9Lectures

Problem Definition, Frequent Item Set Generation, The APRIORI Principle, Support and Confidence Measures, Association Rule Generation; APRIORI Algorithm, The Partition Algorithms, FP-Growth Algorithms, Compact Representation of Frequent Item Set- Maximal Frequent Item Set, Closed Frequent Item Set.

Learning outcomes: At the end of this unit, the students will be able to

- Apply Apriori & FP growth algorithm to generate association rules.
- Represent frequent item sets in a compact way.

UNIT-IV Classification

9 Lectures

Problem Definition, General Approaches to solving a classification problem, Evaluation of Classifiers, Classification techniques, Decision Trees-Decision tree Construction, Methods for Expressing attribute test conditions, Measures for Selecting the Best Split, Algorithm for Decision tree Induction; Naive-Bayes Classifier, Bayesian Belief Networks; K-Nearest Neighbour classification-Algorithm and Characteristics, Prediction: Accuracy and Error measures, Evaluating the accuracy of a classifier or a predictor, Ensemble methods.

Learning outcomes: At the end of this unit, the students will be able to

- Apply classification algorithms such as decision trees, Naïve bayes classifier, Bayesian belief networks and KNN for prediction.
- Analyse classifiers using classification metrics.
- Know the importance of ensemble methods.

UNIT-V Clustering

9Lectures

Clustering Overview, A Categorization of Major Clustering Methods, partitioning methods, hierarchical methods, partitioning clustering-k-means algorithm, Hierarchical clustering-agglomerative methods and divisive methods, Basic Agglomerative Hierarchical Clustering Algorithm, Key Issues in Hierarchical Clustering, Strengths and Weakness, Outlier Detection.

Learning outcomes: At the end of this unit, the students will be able to

- Categorize clustering techniques.
- Apply basic clustering methods to cluster the data.
- Identify outliers in the given data.

TEXTBOOKS

1. Data Mining-Concepts and Techniques- Jiawei Han, Micheline Kamber, JianPei, Morgan Kaufmann Publishers, Elsevier, 3Edition, 2012.

REFERENCES

1. Data Mining Techniques, Arun K Pujari, 3rd Edition, Universities Press.
2. Data Ware Housing Fundamentals, Pualraj Ponnaiah, Wiley Student Edition.
3. The Data Ware House Life Cycle Toolkit-Ralph Kimball, Wiley Student Edition.
4. Data Mining, Vikaram Pudi, P Radha Krishna, Oxford University.

CHANGE OF SYLLABUS

Unit No	Changes Incorporated
Unit-2 Introduction to Data Mining	<u>Topics included</u> • Applications and Trends in Data mining 1% concepts pertaining to applications and trends are included

1.1.3 of NAAC

Name of the Course	Course Code	Year of Introduction	Activities/Content with a direct bearing on Employability/ Entrepreneurship/ Skill development	Mapping with Employability/Skill development/ Entrepreneurship
Data Warehousing and Data Mining	IT314(B)	R20	APRIOIRI Algorithm, K-Nearest Neighbour classification algorithm, FP-Growth Algorithms, Compact Representation of Frequent Item Set.	Employability