

Data Warehousing and Mining & Business Intelligence

B.E. Sem. VII [INFT]

EVALUATION SYSTEM

	Time	Marks
Theory Exam	3 Hrs.	100
Practical Exam	–	–
Oral Exam	–	25
Term Work	–	25

Prerequisite: Data Base Management System.

Objective: Today is the era characterized by Information Overload – Minimum knowledge. Every business must rely extensively on data analysis to increase productivity and survive competition. This course provides a comprehensive introduction to data mining problems concepts with particular emphasis on business intelligence applications.

The three main goals of the course are to enable students to:

1. Approach business problems data-analytically by identifying opportunities to derive business value from data.
2. Know the basics of data mining techniques and how they can be applied to extract relevant business intelligence.

SYLLABUS

1. Introduction to Data Mining

Motivation for Data Mining, Data Mining-Definition & Functionalities, Classification of DM systems, DM task primitives, Integration of a Data Mining system with a Database or a Data Warehouse, Major issues in Data Mining.

2. Data Warehousing – (Overview Only)

Overview of concepts like star schema, fact and dimension tables, OLAP operations, From OLAP to Data Mining.

3. Data Preprocessing

Why? Descriptive Data Summarization, Data Cleaning: Missing Values, Noisy Data, Data Integration and Transformation. Data Reduction:-Data Cube Aggregation, Dimensionality reduction, Data Compression, Numerosity Reduction, Data Discretization and Concept hierarchy generation for numerical and categorical data.

4. Mining Frequent Patterns, Associations, and Correlations

Market Basket Analysis, Frequent Itemsets, Closed Itemsets, and Association Rules, Frequent Pattern Mining, Efficient and Scalable Frequent Itemset Mining Methods, The Apriori Algorithm for finding Frequent Itemsets Using Candidate Generation, Generating Association Rules from Frequent Itemsets, Improving the Efficiency of Apriori, Frequent Itemsets without Candidate Generation using FP Tree, Mining Multilevel Association Rules, Mining Multidimensional Association Rules, From Association Mining to Correlation Analysis, Constraint-Based Association Mining.

5. Classification & Prediction

What is it? Issues regarding Classification and prediction:

- **Classification methods:** Decision tree, Bayesian Classification, Rule based
- **Prediction:** Linear and non linear regression

Accuracy and Error measures, Evaluating the accuracy of a Classifier or Predictor.

6. Cluster Analysis

What is it? Types of Data in cluster analysis, Categories of clustering methods, Partitioning methods – K-Means, K-Medoids. Hierarchical Clustering- Agglomerative and Divisive Clustering, BIRCH and ROCK methods, DBSCAN, Outlier Analysis

7. Mining Stream and Sequence Data

What is stream data? Classification, Clustering Association Mining in stream data. Mining Sequence Patterns in Transactional Databases.

8. Spatial Data and Text Mining

Spatial Data Cube Construction and Spatial OLAP, Mining Spatial Association and Co-location Patterns, Spatial Clustering Methods, Spatial Classification and Spatial Trend Analysis. **Text Mining** Text Data Analysis and Information Retrieval, Dimensionality Reduction for Text, Text Mining Approaches.

9. Web Mining

Web mining introduction, Web Content Mining, Web Structure Mining, Web Usage mining, Automatic Classification of web Documents.

10. Data Mining for Business Intelligence Applications

Data mining for business Applications like Balanced Scorecard, Fraud Detection, Clickstream Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance and CRM etc.

Reference :

1. Data Mining Concepts and Techniques (*Han, Kamber*) Morgan Kaufmann 2nd Edition
2. Introduction to Data Mining (*P. N. Tan, M. Steinbach, Vipin Kumar*) Pearson Education
3. Data Mining with Microsoft SQL Server 2008 (*MacLennan Jamie, Tang ZhaoHui and Crivat Bogdan*) Wiley India Edition.
4. Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner (*G. Shmueli, N.R. Patel, P.C. Bruce*) Wiley India.
5. Data Mining Techniques (*Michael Berry and Gordon Linoff*) 2nd Edition Wiley Publications.
6. Data Mining and Data Warehousing and OLAP (*Alex Berson and Smith*) McGraw Hill Publication.
7. Decision Support and Data Warehouse Systems (*E. G. Mallach*) Tata McGraw Hill.
8. Mastering Data Mining- Art & science of CRM (*Michael Berry and Gordon Linoff*) Wiley Student Edition
9. Multidimensional Data Analysis and Data Mining (*Arijay Chaudhry & P. S. Deshpande*) Dreamtech Press
10. Data Mining (*Vikram Pudi & Radha Krishna*) Oxford Higher Education.
11. Mining the Web: Discovering knowledge from hypertext data (*Chakrabarti, S.*)
12. Fundamentals of Data Warehouses (*M. Jarke, M. Lenzerini, Y. Vassiliou, P. Vassiliadis (ed.)*) Springer-Verlag, 1999.

