Part I

Course Title: Data Warehousing and Data Mining

Course Code: CS4483

Course Duration: One Semester

No. of Credit Units: 3

Level: B4

Medium of Instruction: English

Prerequisites: (Course Code and Title)
CS3402 Database Systems or
CS3462 Introduction to Database Systems

Precursors: (Course Code and Title)
Nil

Equivalent Courses: (Course Code and Title)
Nil

Exclusive Courses: (Course Code and Title)
CS3481 Fundamentals of Data Science

Part II

1. **Course Aims:**

   This course aims to explore a new field and frontier in database systems, "data warehousing and data mining". The principles, algorithms, methodology, and applications are introduced. The components of data warehouse, including data source and transformation tools, metadata management, query reporting and OLAP are discussed. The course also covers techniques and algorithms in data mining, including association rule mining, cluster analysis and data classification.
2. Course Intended Learning Outcomes (CILOs)
(state what the student is expected to be able to do at the end of the course according to a given standard of performance)

Upon successful completion of this course, students should be able to:

<table>
<thead>
<tr>
<th>No.</th>
<th>CILOs</th>
<th>Weighting (if applicable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>identify the main characteristics of different data warehousing and data mining techniques through observation of their operations;</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>perform a critical assessment of current data warehousing and data mining techniques;</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>implement the main algorithms in data warehousing and data mining in a computationally efficient way.</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>propose new solutions for data warehousing and data mining problems by improving and combining current techniques;</td>
<td></td>
</tr>
</tbody>
</table>

3. Teaching and learning Activities (TLAs)
(designed to facilitate students’ achievement of the CILOs)

Teaching pattern:
*Suggested lecture/tutorial/laboratory mix:* 2 hrs. lecture; 1 hr. tutorial

<table>
<thead>
<tr>
<th>ILO No</th>
<th>TLAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>CILO 1 CILO 2</td>
<td>This course will focus on introducing the fundamental and state-of-the-art techniques in data warehousing and data mining. The exact set of topics to be covered may vary from year to year, depending on the trend and direction of this emerging field. For each topic, instructor will introduce background information, fundamental and advanced techniques. Students are then required to have a comprehensive understanding of these techniques by reading related literature and implementing a selected set of algorithms for experiment and performance comparison.</td>
</tr>
<tr>
<td>CILO 1 CILO 2</td>
<td><strong>Assignment</strong> – This assignment gives students an opportunity to identify the characteristics of different techniques in data warehousing and data mining. Students are also required to search literature and perform critical assessment of recent data warehousing and data mining methodologies. This activity supports Course ILOs #1, 2</td>
</tr>
<tr>
<td>CILO 3 CILO 4</td>
<td><strong>Project</strong> – There will be two projects: The first project gives students an opportunity to implement</td>
</tr>
</tbody>
</table>
existing algorithms in data warehousing and data mining in a computationally efficient way. The second project allows students to create new designs for data warehousing and data mining systems. This activity supports Course ILOs #3, 4.

4. **Assessment Tasks/Activities**  
*(designed to assess how well the students achieve the CILOs)*

<table>
<thead>
<tr>
<th>ILO No</th>
<th>Type of assessment tasks/activities</th>
<th>Weighting (if applicable)</th>
<th>Remarks</th>
</tr>
</thead>
</table>
| CILO 1 | • **Coursework**: The quality of students’ assignments will be used to assess this ILO.  
• **Exam**: Final exam will include questions to assess the students’ understanding on various topics of data warehousing and data mining. | | |
| CILO 2 | • **Coursework**: Students are required to perform a critical assessment of different data warehousing and data mining techniques in the assignment. The quality and relevancy of their critiques will be a measure of this ILO.  
• **Exam**: Final exam will include questions to test the students’ capability to perform critical assessment of data warehousing and data mining methods. | | |
| CILO 3 | • **Coursework**: In one of the projects, students will implement existing algorithms for specific problems in data warehousing and data mining. The quality of the implementation, will be the measure of this ILO. | | |
| CILO 4 | • **Coursework**: In one of the projects, students will create new designs for data warehousing and data mining systems. The creativity and the effectiveness of the proposed designs will be the measures of this ILO.  
• **Exam**: Final exam will include case studies to evaluate the students’ capability in proposing new solutions for challenging problems in data warehousing and data mining. | | |

5. **Grading of Student Achievement:** Refer to Grading of Courses in the Academic Regulations and to the Explanatory Notes.

*Examination duration: 2 hours  
Percentage of coursework, examination, etc.: 50% CW; 50% Exam  
Grading pattern: Standard (A+AA–…F)*

For a student to pass the course, at least 30% of the maximum mark for the
examination must be obtained.
A formative assessment will be made on the students’ ability to apply tools and knowledge to different situation. The equal weighting of coursework and examination assessment is due to the emphasis on both the practicality and the theory of data warehousing and data mining.

Part III

Keyword Syllabus

Data integration and transformation, Multi-dimensional data cube, Full cube and iceberg cube, On-line analytical processing (OLAP), Apriori based algorithms, FP-tree, $k$-means, $k$-mediods, BIRCH, DBSCAN, Decision trees, Bayesian classifier, Neural network and support vector machine.

Syllabus

1. Knowledge discovery process

   Introduce KDD (knowledge discovery) process in three stages: data warehousing, data mining, and knowledge presentation. In addition, basic data preparation techniques including data cleaning, selection, integration, transformation and reduction will be discussed. Fundamental data analysis techniques such as attribute relevance analysis, principle component analysis will be emphasized.

2. Data warehousing

   Contrast traditional database and data warehouse, SQL and DMQL, OLTP and OLAP. Introduce multi-dimensional data cube model, OLAP operations, popular schemas (star, snowflake), and the related computational issues in full cube and iceberg cube.

3. Association Mining

   Discuss several algorithms in frequent pattern mining and association rule mining, including Apriori-based algorithms for frequent and sequential pattern mining, FP-tree data structure and FP-growth algorithm. Introduce various constraints including monotonic, anti-monotonic, succinct, convertible constraints for efficient mining.

4. Cluster Analysis

   Introduce the major types of clustering techniques: partitioning, hierarchical, density-based, grid-based, model based clustering. Popular algorithms such as $k$-means, BIRCH and DBSCAN will be discussed. Related issues in outlier analysis and detection will be introduced.
5. Classification Methods

Introduce popular classifiers being used for database and data warehouse. These techniques include decision tree, Bayesian classifier, neural network and support vector machines. Introduce data sampling and model evaluation methods for classification.

Recommended Reading:

Text(s):

Essential Text

Tan P. N. and Steinbach M. and Kumar V. Introduction to Data Mining. Addison Wesley (2005)

Supplementary Reading

Poe V. Building a Data Warehouse for Decision Support. 2nd Ed. Prentice-Hall PTR (1997)


Berson A. and Smith S. Data Warehousing, Data Mining, and OLAP. McGraw-Hill (1997)