CS 536 Data Mining

Instructor's Name: Asim Karim    Year: 2004-05
Office No. & Email: 429, akarim@lums.edu.pk    Quarter: Autumn
Office Hours: TBA    Category: MS
TA for the Course: TBA

<table>
<thead>
<tr>
<th>Course Code (Units)</th>
<th>CS 536 Data Mining (3 Units)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Course Description</strong></td>
<td>Data mining or the discovery of knowledge in large datasets has created a lot of interest in the database and data engineering communities in recent years. The tremendous increase in the generation and collection of data has highlighted the urgent need for systems that can extract useful and actionable knowledge from large datasets. This course will provide a comprehensive introduction to the data mining process; build theoretical and conceptual foundations of key data mining tasks such as association rules mining and clustering; discuss analysis and implementation of algorithms; and introduce major research sub-areas such as web/network mining. Emphasis will be placed on the design and implementation of efficient and scalable algorithms. The students will get hands on experience through the implementation of algorithms in programming exercises. Selected research papers will also be discussed in class to supplement the text book material.</td>
</tr>
</tbody>
</table>

| Core/Elective | This is an elective course. |
| Pre-requisites | CS 213 Data Structures and Algorithms; CS 341 Databases (recommended) Students are expected to have a basic understanding of the terminologies and concepts of database systems and proficiency in algorithm implementation in C/C++/JAVA. |
| Goals | 1. Provide a comprehensive introduction to data mining  
2. Develop conceptual and theoretical understanding of the data mining process  
3. Provide hands-on experience in the implementation and evaluation of data mining algorithms  
4. Develop interest in data mining research |
| TextBooks, Programming Environment, etc. | Textbook:  
1. Data Mining: Concepts and Techniques, Jiawei Han and Micheline Kamber, Morgan Kaufmann Publishers, 2001.  
**Supplementary Materials:**  
4. Selected papers. |
CS 536 Data Mining

Year: 2004-05
Quarter: Autumn

Lectures, Tutorials & Attendance Policy

There will be 20 sessions of 75 minutes each, one in-class midterm exam and one final exam. There are no points for attendance; however, students who miss lectures will find it very difficult to make up for the content covered, with the possible additional penalty of missing quizzes.

Grading

20% Programming exercises and assignments
10% Quizzes
15% Project
25% Midterm Exam
30% Final Exam (Comprehensive)

Additional Details

The course website will be the primary source for announcements and reading material including lecture slides, handouts, and web links. http://suraj.lums.edu.pk/~cs536a04

Cheating and plagiarism will not be tolerated and will be referred to the disciplinary committee for appropriate action. If an assignment is discussed among students, it is required that each student writes up the solution independently, and without looking at notes from the discussion. Downloading code segments from the internet and presenting them as your own is considered plagiarism.

Late Submission Policy:
Programming exercises will not be acceptable after the deadline.
## CS 536 Data Mining

**Year:** 2004-05  
**Quarter:** Autumn

<table>
<thead>
<tr>
<th>Topics</th>
<th>Sessions</th>
<th>Readings</th>
</tr>
</thead>
</table>
| 1. An Overview of Data Mining  
Need and motivation; data mining process; data mining tasks and functionalities, interestingness measures | 1-2 | Ch 1 |
| 2. Data Preprocessing  
Data cleaning, data integration and transformation, data reduction, discretization, concept hierarchies | 3-7 | Ch 3 |
| 3. Mining Association Rules  
Basic definitions, market basket analysis, types of association rules, interestingness measures, frequent Item-set generation, Apriori algorithm, TIDApriori algorithm, Hybrid Apriori algorithm, FP- growth algorithm, mining multilevel, multidimensional and quantitative association rules | 8-12 | Ch 6, Selected Research Papers |

### MIDTERM EXAM 25%

<table>
<thead>
<tr>
<th>Topics</th>
<th>Sessions</th>
<th>Readings</th>
</tr>
</thead>
</table>
| 4. Clustering Analysis  
Basic terminology, Partitioning methods: K-Means, K-Medoids, CALARANS  
Hierarchical methods: BIRCH, CURE etc  
Density based methods: DBSCAN, OPTICS  
Grid based methods: STING, WaveCluster, CLIQUE | 14-17 | Ch 8, Selected Research Papers |
| 5. Mining Complex Data Types  
mixing time-series and sequence datasets; web/network data mining; security and data mining | 18-20 | Ch 9, Selected Research Papers |

### FINAL EXAM 30%

<table>
<thead>
<tr>
<th>Topics</th>
<th>Sessions</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>19-20</td>
<td></td>
<td>19-20</td>
</tr>
</tbody>
</table>