

COURSE DESCRIPTION

Dept., Number	CS422	Course Title	Introduction to Data Mining
Semester hours	3	Course Coordinator	Dr. Nazli Goharian, Clinical Associate Professor

Current Catalog Description

This course will provide an introductory look at concepts and techniques in the field of data mining. After covering the introduction and terminologies to Data Mining, the techniques used to explore the large quantities of data for the discovery of meaningful rules and knowledge such as market basket analysis, nearest neighbor, decision trees, neural networks, and clustering are covered. The students learn the material by implementing different techniques throughout the semester (3-0-3).

Textbook

J. Han, M. Kamber. Data Mining Concepts and Techniques, Morgan Kaufmann, 2006

References

--

Course Outcomes

Students should be able to:

- Explain the Data Mining motivation and applications.
- Explain the Data Mining Architecture.
- Explain Data Preprocessing motivation and techniques.
- Explain various Data Mining algorithms such as Naïve Bayes, Neural Networks, Decision Tree, Association-Rules, and Clustering.
- Explain the scalability issues for each of the algorithms discussed in the class and how they can be modified for scalability.
- Design and implement data mining systems using various data pre-processing techniques and mining algorithms.
- Apply the research ideas into their experiments in building data mining systems.

Relationship between Course Outcomes and Program Outcomes

The following Program Outcomes are supported by the above Course Outcomes:

- a. An ability to apply knowledge of computing and mathematics appropriate to the discipline

- c. An ability to design, implement and evaluate a computer-based system, process, component, or program to meet desired needs
- d. An ability to function effectively on teams to accomplish a common goal
- f. An ability to communicate effectively with a range of audiences
- j. An ability to apply mathematical foundations, algorithmic principles, and computer science theory in the modeling and design of computer-based systems in a way that demonstrates comprehension of the tradeoffs involved in design choices

Prerequisites by Topic

Data Structures, Algorithm and Strong Object Oriented Programming.

Major Topics Covered in the Course

1. Introduction to Data Mining	1.5 hours
2. Data preprocessing	3 hours
3. Classification & Cross Validation	3 hours
4. Evaluation	3 hours
5. Naive Bayes	3 hours
6. Neural Networks	4.5 hours
7. Decision Tree	4.5 hours
8. Rule Based Classification	3 hours
9. K-Nearest Neighbor	3 hours
10. Ensemble Methods	3 hours
11. Association rules	3 hours
12. Cluster analysis	4.5 hours
13. Students Presentations	6 hours
Final Exam	-
	45 hours

Assessment Plan for the Course

End of every semester Course Objective Assessments by CS department. End of semester Course Evaluations by IIT. Reviewed every Spring semester by CS Undergraduate Studies Committee for possible updates in the following Fall. Once every 4-5 years a detailed review of all materials for the course is made by the CS Undergraduate Studies Committee.

How Data in the Course is Used to Assess Program Outcomes (unless adequately covered already in the assessment discussion under Criterion 4)

See the assessment discussion under Criterion 4

For a computer science program

Estimate Curriculum Category Content (Semester hours)

Area	Core	Advanced	Area	Core	Advanced
Algorithms		1	Software design		1
Data structures		1	Concepts of programming languages		