EECE 633: Data Mining

Catalog description:
This course will explore Data Mining theory and examine related applications. Applications include mining of financial data, web data, multimedia and biological data. The course will provide students with the tools for discovering information in large data sets. Data mining refers to knowledge discovery from huge amount of data, and finding non-trivial conclusions. Topics will include methods for preparing data for data mining, statistics, data visualization, business intelligence, knowledge mining, and databases, with a focus on analysis of large data sets. The course will target at least one new data mining problem involving real data, for which the students will have to find a solution.

Credit hours: 3 credits

Required or elective:
Elective for CCE / ECE students

Prerequisites:
By course: EECE 330, or consent of the instructor
By topic: Programming Language (C++/Java), Data Structures, or consent of the instructor
Desirable: Database Management (Highly desirable), Software Engineering, Algorithms, Pattern Recognition, Statistics.

Textbook(s) and/or required materials

References:

Course Objectives
1. Provide students with an overview of the methodologies and approaches to data mining
2. Gain insight into the challenges and limitations of different data mining techniques
3. Provide the students with practice on applying data mining solutions using common data mining software tool (e.g. WEKA, SPSS, Data Miner,...)
4. Prepare students for research in the area of data mining and related applications
5. Enhance students communication and problem solving skills

Course Topics

1. Overview of Data mining: Definition, Challenges, and Applications (Ch. 1)
2. Data mining process: Data preparation/cleaning, task identifications, and data presentation (ch. 2)
3. Data Warehousing (ch. 3 )
4. Relationship between Data Warehousing, On-line Analytical Processing, and Data Mining. (ch. 3, 4)
5. Association Rule mining (ch. 5)
6. Introduction to a data mining software system (e.g. WEKA - Waikato Environment for Knowledge Analysis) (external references and lab material) (In Lab)
7. Association rules - different algorithm types for data mining (ch. 5)
8. Classification/Prediction as applied for data mining (ch. 6)
9. Classification - tree-based approaches (ch. 6)
10. Clustering (ch. 7)
11. Introduction to advanced topics and applications in data mining, e.g. for: Text Mining, Time Series, Data Streams, Multimedia

Course Learning Outcomes

1. Can define what a typical data mining is and what it can be applied for.
2. Understand the difference between Data Warehousing and general databases
3. Can determine the different steps followed in Data mining and pre-processing for Data mining
4. Are able to apply Association Rule mining
5. Are familiar with a data mining software system and use it for solving data mining problems
6. Can describe and apply at least one of the algorithms used for Association rules in data mining
7. Can describe tree-based approaches for classification
8. Can apply at least one of the Classification methods for data mining
9. Can apply at least one of the Prediction methods for data mining
10. Can describe the Clustering basics and approaches
11. Are introduced to challenges in related advanced applications such as data mining for: Text, Time Series, Data Streams, or Multimedia applications
Class/laboratory schedule
a- Two 75-minute lectures per week.
b- Use of computer lab is needed for working on the projects.

Resources of the course
Reference books, online references, and Moodle.

Computer usage
Programming language needed for data mining projects.

Evaluation methods
1- Assignments (10%)
2- Mid-term (30%)
3- Project(s) (30%)
4- Final Exam (30%)

Professional component
Engineering topics: 75%
General education: 5%
Mathematics and basic sciences: 20%

Person(s) who prepared this description and date of preparation
Hazem M. Hajj, May 2009

Date of last revision
February 2013