INSY 5339 – Business Data Mining
Spring 2015

Syllabus

Instructor:
Dr. Riyaz Sikora
Office: 523 COBA
Office Hours: After Class or by appt.
rsikora@uta.edu
817-272-5397

Brief Description:
This course will cover the foundations of business data mining. Data mining is the extraction of implicit, previously unknown, and potentially useful information from data. Data that has relevance for managerial decisions is accumulating at an incredible rate due to a host of technological advances. Electronic data capture has become inexpensive and ubiquitous as a by-product of innovations such as the internet, e-commerce, electronic banking, point-of-sale devices, bar-code readers, and intelligent machines. Such data is often stored in data warehouses and data marts specifically intended for management decision support. Finding patterns, trends, and anomalies in these datasets, and summarizing them with simple quantitative models, is one of the grand challenges of the information age.

This course will examine tools and techniques from the fields of machine learning (AI) and statistics used in practical data mining for finding, and describing, structural patterns in data. We will survey applications and provide an opportunity for hands-on experimentation with algorithms for data mining using easy-to- use software and cases.

Course Text Book:
Witten and Eibe, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 3rd Edition.

Suggested Readings:
Tan, Steinbach, and Kumar, *Data Mining*, Addison-Wesley
Han, Kamber, and Pei, *Data Mining*, Morgan Kaufmann, 3rd Edition

Software:
Weka data mining software in Java. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Grading:

40% Group project (& presentation)
30% Assignments
30% Two Exams (each worth 15%)

Tentative List of Topics:

- Introduction to Data Mining
- Machine Learning and Classification
- Knowledge Representation
- Classification methods:
    - OneR
    - NaiveBayse
    - Decision Trees
    - Rules
    - Regression
    - Nearest Neighbor
- Performance Evaluation
- Data Preparation for Knowledge Discover
- Clustering
- Association Rules

Important Dates:

3/26: Exam #1
4/9, 4/16, 4/23: Project Presentations
4/30: Exam #2
5/14: Final Project Reports due

UNIVERSITY POLICIES

Attendance Policy: Students are required to read and be prepared to discuss the assigned textbook chapters and workbook exercises on the scheduled class days. Class attendance and lateness policies will be discussed during the first week of class.

Drop Policy: Students may drop or swap (adding and dropping a class concurrently) classes through self-service in MyMav from the beginning of the registration period through the late registration period. After the late registration period, students must see their academic advisor to drop a class or withdraw. Undeclared students must see an advisor in the University Advising Center. Drops can continue through a point two-thirds of the way through the term or session. It is the student's responsibility to officially withdraw if they do not plan to attend after registering. **Students will not be automatically dropped for non-attendance**. Repayment of certain types of financial aid administered through the University may be required as the result of dropping classes or withdrawing. Contact the Financial Aid Office for more information.

Americans with Disabilities Act: The University of Texas at Arlington is on record as being committed to both the spirit and letter of all federal equal opportunity legislation, including the *Americans with Disabilities Act (ADA)*. All instructors at UT Arlington are required by law to provide "reasonable accommodations" to students with disabilities, so as not to discriminate on the basis of that disability. Any student requiring an accommodation for this course must provide the instructor with official documentation in the form of a letter certified by the staff in the Office for Students with Disabilities, University Hall 102. Only those students who have officially documented a need for an accommodation will have their request honored. Information regarding diagnostic criteria and policies for obtaining disability-based academic accommodations can be found at www.uta.edu/disability or by calling the Office for Students with Disabilities at (817) 272-3364.

Academic Integrity: It is the philosophy of The University of Texas at Arlington that academic dishonesty is a completely unacceptable mode of conduct and will not be tolerated in any form. All persons involved in academic dishonesty will be disciplined in accordance with University regulations and procedures. Discipline may include suspension or expulsion from the University. According to the UT System Regents' Rule 50101, §2.2, "Scholastic dishonesty includes but is not limited to cheating, plagiarism, collusion, the submission for credit of any work or materials that are attributable in whole or in part to another person, taking an examination for another person, any act designed to give unfair advantage to a student or the attempt to commit such acts."

Student Support Services Available: The University of Texas at Arlington provides a variety of resources and programs designed to help students develop academic skills, deal with personal situations, and better understand concepts and information related to their courses. These resources include tutoring, major-based learning centers, developmental education, advising and mentoring, personal counseling, and federally funded programs. For individualized referrals to resources for any reason, students may contact the Maverick Resource Hotline at 817-272-6107 or visit www.uta.edu/resources for more information.

Electronic Communication Policy: The University of Texas at Arlington has adopted the University "MavMail" address as the sole official means of communication with students. MavMail is used to remind students of important deadlines, advertise events and activities, and permit the University to conduct official transactions exclusively by electronic means. For example, important information concerning registration, financial aid, payment of bills, and

graduation are now sent to students through the MavMail system. All students are assigned a MavMail account. ***Students are responsible for checking their MavMail regularly.*** Information about activating and using MavMail is available at http://www.uta.edu/oit/email/. There is no additional charge to students for using this account, and it remains active even after they graduate from UT Arlington.

To obtain your NetID or for logon assistance, visit https://webapps.uta.edu/oit/selfservice/. If you are unable to resolve your issue from the Self-Service website, contact the Helpdesk at helpdesk@uta.edu.

THE INSTRUCTOR RESERVES THE RIGHT TO MAKE CHANGES TO THE SYLLABUS AS NECESSARY; IT IS THE STUDENT'S RESPONSIBILITY TO BE AWARE OF THESE CHANGES.