**IE 5300 Topics in Industrial Engineering: Data Mining and Analytics**

**Tuesday & Thursday** **11:00AM – 12:20 PM, WH 402**

**Class Syllabus – Spring 2014**

**Instructor:** Dr. Shouyi Wang (Assistant Professor, Industrial and Manufacturing Systems Engineering Department)

**Office Location:** Office 420H – Woolf Hall

**Class Time**: Tuesday & Thursday 11:00 am to 12:20 pm

**Office Hours:** Tuesday & Thursday 1:30 am to 3:00 pm (other times by appointment and by phone/e-mail)

**Phone:** 817-272-2921 **Email:** shouyiw@uta.edu, **Cell**: 732-208-7249 (**text me** during the work day if you really need to contact me, call my cell after working hours.)

**Required Textbook:**  *No textbook is required and class materials will be provided by the instructor. Recommended Books:*

*Pattern Recognition and Machine Learning by C. M. Bishop*

*Pattern Classification by R. O. Duda, P. E. Hart, and D. G. Stork*

*Learning From Data by Y. S. Abu-Mostafa, M. Magdon-Ismail*

**TA Information:**

Xinglong Ju

TA Office Location: 425A Woolf Hall, IMSE Graduate Assistants Office.

**Course Description:** This course provides a broad introduction to data mining, machine learning and statistical pattern recognition. The basic theories, algorithms, key technologies in data analytics will be discussed. Topics include data representation, feature extraction, feature selection, correlation analysis, classification, pattern recognition, supervised learning (parametric/non-parametric algorithms, support vector machines, kernels, neural networks), unsupervised learning (clustering, dimensionality reduction, recommender systems, deep learning), and reinforcement learning. The course will discuss many case studies and real-world applications. You will learn how to process massive data, and apply the most effective data mining and machine learning techniques to solve challenging engineering and scientific problems. You will gain the practical know-how needed to quickly and powerfully apply these techniques to solve data mining and knowledge discovery problems.

**Course Goals/Learning Objectives:** This course will provide students with the insights of the integration of data mining in engineering and operations research. Students will be prepared to implement data mining techniques to extract patterns, trends, and other useful information from ***large volumes of data***, and to solve ***interdisciplinary problems*** for enhanced data-driven decision-making.

**Topics Covered:**

***(1) Data Mining Overview***  
• What is data mining?   
• Overview of data mining tools: classification, regression, clustering, anomaly detection, association rules   
• Applications of data mining

***(2) Data Representation and Processing***   
• Spectra (or time series) data analysis   
• Preprocessing of data and dimensionality reduction   
• Feature extraction and selection with spectra data   
• Wavelets-based approaches   
***(3) Feature Extraction***

• Statistical features, pattern features, frequency features, time-frequency features, spatial-temporal features  
• Metrics and distance functions depending on data types (e.g., Dynamic time warping)  
• Separability measures, distance between probability distributions (e.g., divergence)

***(4) Feature Selection Approaches***

• Statistical variable selection

• Optimization methods for feature selection   
• Branch and bound method   
• Genetic algorithms   
• Current research issues (SVM-based feature selection, Huge-scale feature selection, Wavelet-based feature selection, Pattern-based feature selection

***(5) Classification and Supervised Learning Models***

• Linear Models for Classification

• Support vector machines

• Kernel methods

• Neural networks

• Non-parametric algorithms (KNN)  
***(6) Regression-Based Decision Making***  
• Loss functions   
• Regularization (ridge regression, Lasso, and others)   
• Variable selection in regression (vs feature selection in classification)   
• Support vector machine for regression   
• Function approximation

***(7) Bayesian Decision Making***

• Bayesian decision theory

• Maximum-likelihood and Bayesian parameter estimation  
• Hidden Markov Models

***(8) Unsupervised Learning***

• Clustering and dimensionality reduction

• Unsupervised Bayesian learning and classification

• Principle Component Analysis (PCA)

• Independent Component Analysis (ICA)

• Association Rules and Recommender systems

• Reinforcement Learning

***(9) Data-Driven Decision-Making Systems***  
• Integration of data mining and machine learning techniques to decision making

• Case studies in medical, industry, financial, engineering, and scientific applications

**Attendance and Drop Policy:** You are responsible for the material presented in class when you are absent. You ***must*** provide advanced notice if you will be out for an exam. You may drop this course through the University published last day to drop. Attendance won’t be taken, but the exams will be based on the lectures and materials presented.

**Key Schedule Dates:**

|  |  |
| --- | --- |
| Mar 6th | Mid-Term Exam |
| May 1st | Project Presentation |
| May 5th | Final Project Report Due |
| May 8th | Final Exam |

**Course Evaluation & Final Grade:**

* Homework and Quizzes counting 50% of total grade.
* Project and Presentation: 50% of total grade. This must be completed in order to receive a grade in the class. It is a “key assignment”. Not submitting the project will result in an ‘incomplete’ grade for the course.
* Late homework/project: Once the class starts, late homework or project assignments that come in during the class will have a minimum of 20% docked. Late homework will not be accepted after the next class period past the due date. A late project, past the due date, will lose a minimum of 10% points.

**General Policies/Guidelines:**

* If you require an accommodation based on disability, please meet with me in the privacy of my office, during the first week of the semester, to make sure you are properly accommodated.
* All tests and quizzes are open book, closed notes (1 formula sheet 8.5 x 11” is allowed for tests). The formula sheet may not include worked problems. **You must turn in your formula sheet with your exam**.
* Students must work alone on tests and quizzes. No sharing of any materials may occur during tests and quizzes including books, calculators, formula sheets, etc.
* Students need to individually turn in homework assignments, e.g. one assignment turned in per student with your name on it. However, homework can be worked jointly with other students.
* If you miss an exam, **you must have an approved** excuse (i.e. from a medical doctor with contact information indicating you were too ill to sit for an exam or have received permission from the instructor at least a couple of days in advance of the exam).
* Cell Phones / pagers / etc. should be turned off or set to vibrate during class. Any necessary calls and communication should be made outside the class. Please be courteous to the instructor and your fellow classmates.

**Key Assignment:**

During the semester the one major data mining project assignment will be designated as a **key assignment**. To pass this class you must successfully complete all key assignment requirements. Key assignments and their respective grading criteria will be announced in each particular course. Late **key assignments** will be accepted for completing the class but may not be accepted for the course grade depending on the instructor policy. In summary, if you do not successfully complete the **key assignment**, you will receive an incomplete (X) in the class.

**Electronic Communication:**

UT Arlington has adopted MavMail as its official means to communicate with students about important deadlines and events, as well as to transact university-related business regarding financial aid, tuition, grades, graduation, etc. All students are assigned a MavMail account and are responsible for checking the inbox regularly. There is no additional charge to students for using this account, which remains active even after graduation. Information about activating and using MavMail is available at <http://www.uta.edu/oit/cs/email/mavmail.php>.

**Student Feedback Survey:**

At the end of each term, students enrolled in the class shall be directed to complete an online Student Feedback Survey (SFS). Instructions on how to access the SFS for this course will be sent directly to each student through MavMail approximately 10 days before the end of the term. Each student’s feedback enters the SFS database anonymously and is aggregated with that of other students enrolled in the course. UT Arlington’s effort to solicit, gather, tabulate, and publish student feedback is required by state law; students are strongly urged to participate. For more information, visit <http://www.uta.edu/sfs>. Please fill in the on-line evaluation. We do pay attention to your feedback!

**Final Review Week:**

A period of five class days prior to the first day of final examinations in the long sessions shall be designated as Final Review Week. The purpose of this week is to allow students sufficient time to prepare for final examinations. During this week, there shall be no scheduled activities such as required field trips or performances; and no instructor shall assign any themes, research problems or exercises of similar scope that have a completion date during or following this week *unless specified in the class syllabus*. During Final Review Week, an instructor shall not give any examinations constituting 10% or more of the final grade, except makeup tests and laboratory examinations. In addition, no instructor shall give any portion of the final examination during Final Review Week. During this week, classes are held as scheduled. In addition, instructors are not required to limit content to topics that have been previously covered; they may introduce new concepts as appropriate.

**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](http://www.uta.edu/disability) or by calling the Office for Students with Disabilities at (817) 272-3364.

**Academic Dishonesty:**

Students enrolled in this course are expected to adhere to the UT Arlington Honor Code:

*I pledge, on my honor, to uphold UT Arlington’s tradition of academic integrity, a tradition that values hard work and honest effort in the pursuit of academic excellence.*

*I promise that I will submit only work that I personally create or contribute to group collaborations, and I will appropriately reference any work from other sources. I will follow the highest standards of integrity and uphold the spirit of the Honor Code.*

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.

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. (Regents Rules and Regulations, Part One, Chapter VI, Section 3, Subsection 3.2, Subdivision 3.22).

**Inclement Weather Policy:**

If the University is closed, this class will not meet. Any scheduled assignments or examinations will be rescheduled to the next class period that the class meets. You can get information by dialing 817-272-8821 or 972-601-2049.

**UTA Student Support:**

UT 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. Resources include tutoring, major-based learning centers, developmental education, advising and mentoring, personal counseling, and federally funded programs. For individualized referrals, students may visit the reception desk at University College (Ransom Hall), call the Maverick Resource Hotline at 817-272-6107, send a message to [resources@uta.edu](mailto:resources@uta.edu), or view the information at [www.uta.edu/resources](http://www.uta.edu/resources).

**Emergency Exit Procedures:** Should we experience an emergency event that requires us to vacate the building, students should exit the room and move toward the nearest exit. When exiting the building during an emergency, one should never take an elevator but should use the stairwells. Faculty members and instructional staff will assist students in selecting the safest route for evacuation and will make arrangements to assist handicapped individuals.