

IE 5300-001 Topics in Industrial Engineering: Data Mining and Analytics
Tuesday & Thursday 12:30PM – 1:50 PM, WH 402
Class Syllabus – Spring 2015

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

Office Location: Office 420H – Woolf Hall

Office Telephone Number: 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.)

Faculty Profile: <http://www.uta.edu/profiles/shouyi-wang>

Office Hours: Tuesday & Thursday 2:00 pm to 4:00 pm (other times by appointment and by phone/e-mail)

Section Information: IE 5300 – 001

Class Time & Place: Tuesday & Thursday 12:30 pm to 1:50 pm, WH 402.

TA Information:

Rahil Hosseini, rahil.s.hosseini@gmail.com,

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

Prerequisites: Students entering the class with a pre-existing working knowledge of probability, statistics, algorithms, and computer programming will be at an advantage. Nevertheless, the class has been designed so that anyone with a strong numerate background can catch up and fully participate in this class.

Description of Course Content: 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.

Student Learning Objectives/Outcomes: 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.

Required Textbook and Other Course Materials: *No textbook is required and the instructor will provide class materials. The class mainly uses two books:*

Introduction to Data Mining by Pang-Ning Tan, Michael Steinbach, Vipin Kumar. (The first two topics)

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

Recommended Books.

Pattern Recognition and Machine Learning by C. M. Bishop

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

Descriptions of major assignments and examinations:

- **Assignments:** there will be assignments involving short answers, running existing code on provided datasets, mathematical derivations, and writing some code and pseudocode. All programming assignments will be done using Matlab. The students can also use or practice other programming languages, such as Python, R, that are efficient for data mining and data analysis.

- **Project and Presentation:** once you have seen many examples of data mining methods, you will apply the learned methods in a practical data mining study, and implement the algorithms and ideas using Matlab (or Python/R, etc.). You'll demonstrate that how the data mining methods work and your selected datasets and provide experimental results, and briefly present your work at the end of the class. The students are also encouraged to explore open research questions and data mining problems, a really good job on this part could result in a paper in the end of the class.

Attendance: At The University of Texas at Arlington, taking attendance is not required. Rather, each faculty member is free to develop his or her own methods of evaluating students' academic performance, which includes establishing course-specific policies on attendance. As the instructor of this section, I have decided that attendance at class meetings is not required but strongly encouraged. I expect students to participate actively in the class. If an emergency arises, contact professor prior to class time via email or phone.

Course Evaluation & Grading:

- Homework assignments 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 one week after past the due date. A late project, past the due date, will lose a minimum of 10% points.

Class Policies: Homework assignments will be done individually: each student must hand in their own answers. It **is** acceptable, however, for students to collaborate in helping each other to solve the problems. We will be assuming that, as participants in a graduate course, you will be taking the responsibility to make sure you personally understand the solution to any work arising from such collaboration. Working together is encouraged, **BUT** all programming and analysis are to be done all by yourself. A good rule of thumb is that you should never copy anything other than class notes from another student.

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. For more information, contact the Office of Financial Aid and Scholarships (<http://web.uta.edu/aao/fao/>).

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.

Title IX: The University of Texas at Arlington is committed to upholding U.S. Federal Law "Title IX" such that no member of the UT Arlington community shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity. For more information, visit www.uta.edu/titleIX.

Academic Integrity: Students enrolled all UT Arlington courses 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.

UT Arlington faculty members may employ the Honor Code as they see fit in their courses, including (but not limited to) having students acknowledge the honor code as part of an examination or requiring students to incorporate the honor code into any work submitted. Per UT System *Regents' Rule* 50101, §2.2, suspected violations of university's standards for academic integrity (including the Honor Code) will be referred to the Office of Student Conduct. Violators will be disciplined in accordance with University policy, which may result in the student's suspension or expulsion from the University.

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 classes categorized as “lecture,” “seminar,” or “laboratory” 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>.

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.

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 (stairs), which is located about 10 feet on the right of the door of the classroom WH402. 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 individuals with disabilities.

The English Writing Center (411LIBR): [Optional.] Hours are 9 am to 8 pm Mondays-Thursdays, 9 am to 3 pm Fridays and Noon to 5 pm Saturdays and Sundays. Walk In Quick Hits sessions during all open hours Mon-Thurs. Register and make appointments online at <http://uta.mywconline.com>. Classroom Visits, Workshops, and advanced services for graduate students and faculty are also available. Please see www.uta.edu/owl for detailed information.

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, or view the information at www.uta.edu/resources.

Course Schedule

Class	Topic	Remark
Week 1 01/20, 01/22	Intro to Data Mining, Data Analysis, Probability and Statistics	Introduction and Preliminaries
Week 2 01/27, 01/29	Statistical Concepts and Methods: Evaluating Hypotheses, Sampling, Prediction, Prediction Error, etc.	
Week 3 02/03, 02/05	Data Mining Basic Concepts Visualize and Explore Data Example by Iris Data set SetDaExample: Iris Data	Explore and Visualize Different Types of Data
Week 4 02/10, 02/12	Bayesian Learning and Methods: Bayes Theorem, Gibbs Algorithm, Naïve Bayes Classifier, Bayesian Belief Networks,	Bayes Decision Theory
Week 5 02/17, 02/19	Expectation-Maximization, Hidden Markov Models, Markov Decision Processes, Maximum Likelihood and Least Squared Error	Maximum Likelihood
Week 6 02/24, 02/26	K-nearest Neighbors, Gaussian Mixture Models, Learning, Metrics and Distance, Parzen Windows	Nonparametric Techniques
Week 7 03/03, 03/05	Support Vector Machines and Kernel Methods, Discriminant Analysis for Classification	Discriminant Functions & Analysis
Week 8 03/10, 03/12	Spring Vacation No Lecture	
Week 9 03/17, 03/19	Multilayer Feedforward, Back-propagation, Radial Basis Functions, Deep Learning	Neural Networks Approaches
Week 10 03/24, 03/26	Decision Trees and Information Gain: CART – Classification and Regression Trees	Non-metric Methods
Week 11 03/31, 04/02	Feature Selection Methods in Classification Variable Selection in Regression	Variable Selection Model Formulation
Week 12 04/07, 04/09	Principal Component Analysis, Independent Component Analysis, Multi-Dimensional Scaling	Dimensionality Reduction
Week 13 04/14, 04/16	Clustering Analysis and Unsupervised Learning: K-mean, K-median, hierarchical Clustering, optimization approaches to cluster Analysis	Unsupervised Learning
Week 15 04/21, 04/23	Bagging, Boosting, Cross-Validation, Bootstrapping, Jackknife, VC Dimension	Algorithm Independent Machine Learning
Week 16 04/28, 04/30	Data Mining Case Studies	Comprehensive Case Study & Demo
Week 17 05/05, 05/07	Final Project Presentation	Key Assignment

“As the instructor for this course, I reserve the right to adjust this schedule in any way that serves the educational needs of the students enrolled in this course. –Shouyi Wang.”

Emergency Phone Numbers: In case of an on-campus emergency, call the UT Arlington Police Department at **817-272-3003** (non-campus phone), **2-3003** (campus phone). You may also dial 911.