Instructor Information

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

Office Location: Office 420H – Woolf Hall

Office Telephone Number: 817-272-2921, Cell: 732-208-7249 (text me during the work day if you really need to contact me, call my cell after working hours.)

Email: shouyiw@uta.edu

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

Office Hours: Monday & Wednesday 1:00 PM - 2:30 PM (other times by appointment and by e-mail)
Instructor Office Hours Location: On Teams "IE6318/4314 -Spring 2021 All Sessions"

GTA Information:
Linh Ho Manh, Email: linh.homanh@mavs.uta.edu
GTA Office Hours: Tuesday & Thursday, TBD
GTA Office Hours Location: On Teams "IE6318/4314 -Spring 2021 All Sessions"

Course Information

Section Information:
IE 6318 – 001/002/003/900 & IE 4314-001

Time & Place of Class Meetings:
Monday & Wednesday 2:30 PM to 3:50 PM (on Teams)

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 an in-depth introduction to data mining and 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 algorithm independent machine learning models. 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 Outcomes: This course has three major objectives. First, to provide students with a sound basis in data mining and pattern recognition tasks and techniques. Second, to ensure that students are able to read, and critically evaluate data mining and analytics research papers. Third, to ensure that students are able to implement and to use important data mining and pattern recognition models and algorithms, and to solve interdisciplinary problems for data-driven decision-making problems in engineering and sciences.
Technology Requirements: Students will need to access course content via Canvas, Teams, and Echo360. Students are advised to preview slides/handouts prior to online lectures to help learning performance in the class. Assignments will be submitted electronically on Canvas.

Hybrid Format of the Class: This course will be entirely online, EXCEPT for the final project evaluation/presentations that may take place in-person on campus, complying with university requirements for social distancing, face-coverings, and personal hygiene. Each online lecture recording will be posted on Canvas. We encourage everyone to have active interactions with the instructor in online lectures and office hours to have productive learning for data mining and machine learning concepts and methodologies.

Mandatory Face Covering Policy: All students and instructional staff are required to wear facial coverings while they are on campus, inside buildings and classrooms. Students that fail to comply with the facial covering requirement will be asked to leave the class session. If students need masks, they may obtain them at the Central Library, the E.H. Hereford University Center’s front desk or in their department. Students who refuse to wear a facial covering in class will be asked to leave the session by the instructor, and, if the student refuses to leave, they may be reported to UTA’s Office of Student Conduct.

Required Textbooks and Other Course Materials:
No textbook is required and the instructor will provide class materials. The class mainly follows two books:
Pang-Ning Tan, Michael Steinbach, Vipin Kumar. Introduction to Data Mining. Addison-Wesley, 2005. (The first 3-4 weeks)

Other Recommended Reading Books:

Descriptions of Major Assignments and Examinations:
- Assignments: there will be assignments involving short answers, running existing codes on provided datasets, mathematical derivations, and writing codes and pseudo codes for data mining tasks. We recommend the class to use either Matlab or Python to complete the programming assignments. We will support and provide HW solutions in both Matlab and Python.
- 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. You'll demonstrate 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. The data mining project may result in a conference or journal paper in the end of the class or after 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 course, I have decided that attendance at class meetings is very important and strongly encouraged. I expect our students to attend online lectures and office hours actively to interact with the instructor and solve study problems in time. If an emergency arises, contact professor prior to class time via email or phone.

While UT Arlington does not require instructors to take attendance in their courses, the U.S. Department of Education requires that the University have a mechanism in place to mark when Federal Student Aid recipients “begin attendance in a course.” UT Arlington instructors will report when students begin attendance in a course as part of the final grading process. Specifically, when assigning a student, a grade of F, faculty report the last date a student attended their class based on evidence such as a test, participation in a class project or presentation, or an engagement online via Blackboard. This date is reported to the Department of Education for federal financial aid recipients. Thus, we will take attendance after the census date to record student participation of the class.
Grading Information

Course Evaluation & Grading:
- Homework assignments and quizzes counting 47% of the total grade. Homework will be posted on Canvas under Assignments and should be submitted via Canvas.
- Data mining projects and presentation: 50% of the 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 submission of 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 the due date. A late project, past the due date, will lose a minimum of 20% points.
- Class attendance is very important to facilitate learning and improve efficiency for data mining & machine learning study. We encourage all our students to participate actively in class. Random attendance will be taken over the semester, and the attendance will take 3% of the final score.

Class Policies: Homework assignments will be done individually: each student must hand in their own answers. It is acceptable, however, for students to help each other to debug and solve programming problems/issues. All assignments and data mining analysis should be done by yourself. A good rule of thumb is that you should never copy anything other than class notes from another student. Any form of copying will have severe consequences.

Expectations for Out-of-Class Study:
Beyond the time required to attend each class meeting, students enrolled in this course should expect to spend at least an additional 9 hours per week of their own time in course-related activities, including reading related materials, completing assignments, programming training and practices, etc.

Course Schedule

<table>
<thead>
<tr>
<th>Class</th>
<th>Topic</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Intro to Data Mining, Machine Learning,</td>
<td>Data Mining Introduction</td>
</tr>
<tr>
<td></td>
<td>Real Life Examples</td>
<td></td>
</tr>
<tr>
<td>Week 2</td>
<td>Data Mining Basic Concepts</td>
<td>Data Mining Basic Concepts</td>
</tr>
<tr>
<td></td>
<td>Data Preprocessing Methods</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Principle Component Analysis</td>
<td></td>
</tr>
<tr>
<td>Week 3</td>
<td>Data Similarity/Dissimilarity</td>
<td>Data Exploration &amp; Visualization</td>
</tr>
<tr>
<td></td>
<td>Distance Measures</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Visualize and Explore Data</td>
<td></td>
</tr>
<tr>
<td>Week 4</td>
<td>Classification Basic Concepts</td>
<td>Classification Basics</td>
</tr>
<tr>
<td></td>
<td>Nearest Neighbor Classifier</td>
<td></td>
</tr>
<tr>
<td>Week 5</td>
<td>Decision Trees</td>
<td>Classification Basics</td>
</tr>
<tr>
<td></td>
<td>Model Evaluation Methods</td>
<td></td>
</tr>
<tr>
<td>Week 6</td>
<td>Bayesian Learning and Methods</td>
<td>Bayes Decision Theory</td>
</tr>
<tr>
<td></td>
<td>Bayes Theorem, Naïve Bayes Classifier,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discriminant Functions for Classification</td>
<td></td>
</tr>
<tr>
<td>Week 7</td>
<td>Linear Discriminant Functions for</td>
<td>Linear Discriminant</td>
</tr>
<tr>
<td></td>
<td>Classification</td>
<td>Classification</td>
</tr>
<tr>
<td></td>
<td>Gradient Descent Approaches</td>
<td></td>
</tr>
<tr>
<td>Week 8</td>
<td>Multilayer Neural Networks</td>
<td>Neural Networks</td>
</tr>
<tr>
<td></td>
<td>Back-propagation, Practical Applications</td>
<td></td>
</tr>
<tr>
<td>Week 9</td>
<td>Support Vector Machines</td>
<td>Linear SVM</td>
</tr>
<tr>
<td></td>
<td>Linear SVM</td>
<td></td>
</tr>
<tr>
<td>Week 10</td>
<td>Support Vector Machines</td>
<td>Nonlinear SVM</td>
</tr>
<tr>
<td></td>
<td>Nonlinear SVM, Multiclass SVM</td>
<td></td>
</tr>
<tr>
<td>Week 11</td>
<td>Bagging and Boosting</td>
<td>Ensemble Methods</td>
</tr>
</tbody>
</table>
AdaBoost, Random Forest

<table>
<thead>
<tr>
<th>Week 12</th>
<th>Feature Selection Methods in Classification</th>
<th>Variable Selection in Regression</th>
<th>Variable Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 13</td>
<td>Clustering Analysis and Unsupervised Learning</td>
<td>Unsupervised Learning</td>
<td></td>
</tr>
<tr>
<td>Week 14</td>
<td>Basics of Deep Neural Networks Deep Learning Methods</td>
<td>Deep Learning</td>
<td></td>
</tr>
<tr>
<td>Week 15</td>
<td>Final Project Discussion &amp; Presentation</td>
<td>Project</td>
<td></td>
</tr>
</tbody>
</table>

“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.”

**Institution Information**

UTA students are encouraged to review the below institutional policies and informational sections and reach out to the specific office with any questions. To view this institutional information, please visit the Institutional Information page (http://www.uta.edu/provost/administrative-forms/course-syllabus/syllabus-institutional-policies.php) which includes the following policies among others:

**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://wweb.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 Exam Schedule: This course does not have a final written exam. Each student will complete a data mining project and will make a presentation of the project at the last week of the semester.

Additional Information

Emergency Exit Procedures: Should we experience an emergency event that requires evacuation of the building, students should exit the room and move toward the nearest exit, which is located in the back of the classroom. 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.

Students should also be encouraged to subscribe to the MavAlert system that will send information in case of an emergency to their cell phones or email accounts. Anyone can subscribe at Emergency Communication System.

Student Success Programs
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 by appointment, drop-in tutoring, etutoring, supplemental instruction, mentoring (time management, study skills, etc.), success coaching, TRIO Student Support Services, and student success workshops. For additional information, please email resources@uta.edu, or view the Maverick Resources website.

The IDEAS Center (https://www.uta.edu/ideas/) (2nd Floor of Central Library) offers FREE tutoring and mentoring to all students with a focus on transfer students, sophomores, veterans and others undergoing a transition to UT Arlington. Students can drop in or check the schedule of available peer tutors at www.uta.edu/IDEAS, or call (817) 272-6593.

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.

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. Non-emergency number 817-272-3381.

Library Information

Research or General Library Help
Ask for Help
- Academic Plaza Consultation Services (library.uta.edu/academic-plaza)
- Ask Us (ask.uta.edu/)
- Research Coaches (http://libguides.uta.edu/researchcoach)

Resources
- Library Tutorials (library.uta.edu/how-to)
- Subject and Course Research Guides (libguides.uta.edu)
- Librarians by Subject (library.uta.edu/subject-librarians)
- A to Z List of Library Databases (libguides.uta.edu/az.php)
- Course Reserves (https://uta.summon.serialssolutions.com/#/course_reserves)
- Study Room Reservations (openroom.uta.edu/)