

**CSE 4334 Data Mining
Fall 2019, GS 104, Section 001**

Instructor(s): Prof. Sharma Chakravarthy

Office Number: 632 ERB

Office Telephone Number: None

Email Address: sharmac@cse.uta.edu

Office Hours: Tu/Th: 11:00 am to Noon + by appointment

GTA : Anish Rai, email: anish.rai@mavs.uta.edu

Office Hours: Monday/Wednesday 10am to noon + by appointment, ERB 514

Section Information: CSE 5334-001

Time and Place of Class Meetings: Tu/Th 9:30am to 10:50am, GS 104

Course URL: itlab.uta.edu/courses (for syllabus, class schedule, and Lecture notes)

Research URL: itlab.uta.edu/sharma (for MS/PhD thesis research projects)

CANVASSURL: uta.instructure.com/ (for projects, grades, discussion)

Course Description: This is an introductory course on data mining. Data Mining refers to the process of automatic discovery of patterns and knowledge from large data repositories, including databases, data warehouses, Web, document collections, and data streams. We will study the basic topics of data mining, including data preprocessing, overview of data warehousing and OLAP, frequent pattern and association rule mining, correlation analysis, classification and prediction, and clustering, as well as advanced topics covering techniques and applications of data mining as part of big data analysis.

Student Learning Outcomes: A solid understanding of the basic concepts, principles, and techniques in data mining; an ability to analyze real-world applications, to model data mining problems, and to assess different solutions; an ability to design, implement, and evaluate data mining software.

Prerequisites

- For CSE 4334: CSE 3330 Database Systems I and IE 3301 Engineering Probability (or MATH 3313 Introduction to Probability) or consent of instructor.
- For CSE 5334: There is no official prerequisites. You should have sound CSE background from your Bachelor's program (e.g., programming, data structures and algorithms, discrete mathematics, basics of probabilities and statistics). If you don't have database course from anywhere, you are allowed to take the course, but please get the consent of the instructor. You also must get the consent of the instructor if you have CSE deficiency courses to take.

Course Outline: This course will cover various topics under the umbrella of data mining. Includes projects either for developing algorithms or for analyzing given data sets using specified mining techniques. Topics will be covered under the following modules:

Module I: Intro to mining; what is not mining, overview of probability, supervised and unsupervised learning, predictive modeling (decision trees, Naïve Bayes, and may be SVM)

Module II: Cluster analysis (k-means, DBSCAN)

Module III: Association rules (Apriori and FP tree)

Module IV: Outlier detection, Graph-based approaches to mining, Multilayer Network analysis

Required Textbooks and Other Course Materials:

- (Required) Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar. Introduction to Data Mining, 2nd Edition, Pearson, 2019. ISBN-13: 978-0-13-312890-1. (Sample chapters at <http://www-users.cs.umn.edu/~kumar/dmbook/index.php>). Also, available as eBook.
- (Required for relevant chapters) [MRS] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze. Introduction to Information Retrieval, Cambridge University Press. 2008. (Free book at <http://nlp.stanford.edu/IR-book/>)
- (Reference) Jure Leskovec, Anand Rajaraman and Jeff Ullman. Mining of Massive Datasets, 2nd ed., Cambridge University Press. (Free book at <http://www.mmds.org/#ver21>)

Descriptions of major assignments and examinations: Project: Since the emphasis of this course is on learning mining techniques and acquiring a good understanding of its application to different data sets (that you are likely to encounter in real world), there will be a number of hands-on implementation/analysis projects as part of this course. I plan on using R and R studio (Open source license at <https://www.rstudio.com/products/rstudio/download/>). A wikibook on R is also available at <https://www.webpages.uidaho.edu/~stevel/517/Data%20Mining%20Algorithms%20In%20R.pdf>

Students may be asked to present the details of their implementation and/or analysis in the class. Demo and evaluation of the project will have an optional discussion session with the students to discuss the approach taken. Demos are expected for all projects.

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. I will take attendance in the class aperiodically. If you are serious about learning and doing well in the course, you should **not only attend lectures but participate** during the lectures by asking questions in the class. The class presentation on your projects will constitute 5% of overall grade

Grading: There will be 3 hands-on projects that will constitute approximately 45% of the grade. There will be 3 in-class tests that will constitute approximately 50% of the grade. This is an initial proposal. Class attendance and presentation will carry 5% of the grade. The instructor reserves the right to re-distribute the percentages if deemed necessary. The students may be asked to make an in-class presentation on the project experiences. Attendance and class participation is important for doing well on the course. **Grading for undergraduates will be done separately from graduates in a combined offering. No pre-defined cut-offs. Based on past observations, typically, class average corresponds to a B grade. One standard deviation**

above the class average is guaranteed to earn an A; passing grade is 50%. Note that you have to do consistently well on all projects and exams to earn an A grade.

Expectations for Out-of-Class Study: Beyond the time required to prepare for and attend each class meeting, students enrolled in this course should expect to spend at least an additional 6 hours per week in course-related activities, including reading required materials, completing assignments, preparing for exams, etc. Meet the instructor or the TA for any doubts on projects or class lecture material.

Assignments and Deadlines:

- All the assignments must be submitted through Canvass. We will NOT take hardcopy or email submission, unless the university verifies that Canvass was malfunctioning or unavailable. If you are not able to submit through Canvass due to its technical failure, you can email your assignment to us, together with a screenshot showing the technical failure. We will verify with the university.
- Everything is due by 11:59pm on the due date. The deadline is automatically managed by Canvass. You can still turn in assignment after the deadline. However, you automatically lose designated points as specified in the project assignment after the due time, till you reach 0. (Each individual assignment is 100 points.) We cannot waive the penalty, unless there was a case of illness or other substantial impediment beyond your control, with proof in documents. Please come and discuss with me or the TA.
- **Regrading:** Once the grade of a quiz/exam/project is distributed, you will have 5 business days to dispute it and get it re-evaluated. No re-evaluation will be entertained after the 5 day period. For projects, as part of the document, what has been designed and implemented by each partner (if it is done as a team) should be clearly stated and documented. **All team members will get the same grade/penalty on the project.**

Class schedule: Note that class schedule, test, and project due dates are tentative. Project deadlines and test dates may be changed (with notice) based on the progress made in the class. The course officially ends on the day of the final exam.

Make-up Exams: No makeup tests or exams will be given unless there is a justifiable, documented reason.

How to Do Well in This Course: Based on the feedback I have received over the years, Students who get the most out of this course will be the ones who keep up with the material covered in the class on a regular basis. If you want to do well, attend all the lectures, read the assigned sections of the book/papers, and start early on your projects. Working out the assigned sample questions and questions from book chapters will immensely help in doing well on quizzes/tests. If you are having difficulty, you owe it to yourself to get help. **We will be more than happy to help you. Don't be lazy or afraid to come and see us. We will hold extensive office hours. If you can't make it to office hours but really need help, contact one of us for an appointment. I sincerely want all of you to do well. It is your responsibility to check the web site at least twice a week.**

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/>).

Disability Accommodations: UT 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)*, *The Americans with Disabilities Amendments Act (ADAAA)*, and *Section 504 of the Rehabilitation Act*. 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 disability. Students are responsible for providing the instructor with official notification in the form of a letter certified by the **Office for Students with Disabilities (OSD)**. Students experiencing a range of conditions (Physical, Learning, Chronic Health, Mental Health, and Sensory) that may cause diminished academic performance or other barriers to learning may seek services and/or accommodations by contacting:

The Office for Students with Disabilities, (OSD) www.uta.edu/disability or calling 817-272-3364.
Counseling and Psychological Services, (CAPS) www.uta.edu/caps/ or calling 817-272-3671.

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.

If anyone is caught for cheating, plagiarism or collusion either on the project or on the quiz/exam, the grade for the course will be an automatic Fail grade (F). For projects, all members of the team will receive the same grade/penalty; so choose your partner wisely and make sure you are aware of what your partner is doing!

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. 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.

Student Support Services: 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

A detailed course schedule is provided on the first day of class and it is also posted on the course web site.

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.

[We strongly recommend that you place this information at the very end of your course syllabus or in the footer of the first page. We further recommend that you enter the UTA Police Department's emergency phone number into your own mobile phone. For non-emergencies, contact the UTA PD at 817-272-3381.]

This final section is not part of the syllabus template, but a message from the UT Arlington Library.

Faculty members should feel free to incorporate any of the following information into your course syllabus or other course materials.

Library Home Page	http://www.uta.edu/library
Subject Guides	http://libguides.uta.edu
Subject Librarians.....	http://www.uta.edu/library/help/subject-librarians.php
Database List	http://www.uta.edu/library/databases/index.php
Course Reserves.....	http://pulse.uta.edu/vwebv/enterCourseReserve.do
Library Tutorials	http://www.uta.edu/library/help/tutorials.php
Connecting from Off- Campus	http://libguides.uta.edu/offcampus
Ask A Librarian	http://ask.uta.edu

The following URL houses a page where we have gathered many commonly used resources needed by students in online courses: <http://www.uta.edu/library/services/distance.php>.

The subject librarian for your area can work with you to build a customized course page to support your class if you wish. For examples, visit <http://libguides.uta.edu/os> and <http://libguides.uta.edu/pols2311fm> . If you have any questions, please feel free to contact Suzanne Beckett, at sbeckett@uta.edu or at 817.272.0923.