

Course number, Course CRN, COURSE TITLE
CIS 4303 – 80139 – Data Mining
Fall 2019
Texas A&M University-Central Texas

COURSE DATES, MODALITY, AND LOCATION

Course meeting dates: August 26, 2019 – Dec 13, 2019

Class meeting building and room number: Founders Hall 307

Class meeting day and time: Mon Wed 2:30 PM to 3:45 PM

INSTRUCTOR AND CONTACT INFORMATION

Instructor: Rahul Dwivedi

Office: Founders Hall 323N

Phone: 254 – 519 – 5784

Email: rahul.dwivedi@tamuct.edu (preferred) or Canvas inbox.

Office Hours

Mon Wed 12:00 PM – 2:00 PM

Tue Thurs 3:00 PM – 4:00 PM

At other times: By appointment through email

Mode of instruction and course access:

This a face-to-face course that makes extensive use of the TAMUCT Canvas Learning Management System (<https://tamuct.instructure.com>). The course syllabus, schedule, supplemental readings, class announcements, power point slides, learning modules, homework assignments, exams and other course related documents will be posted on Canvas. Each student is responsible for the posted material and should check Canvas several times a week for updates.

Student-instructor interaction:

I typically respond to Canvas email within 24 hours except on weekends. Email is the best mode of communication (avoid phone calls or voice messages).

WARRIOR SHIELD

Emergency Warning System for Texas A&M University-Central Texas

Warrior Shield is an emergency notification service that gives Texas A&M University-Central Texas the ability to communicate health and safety emergency information quickly via email, text message, and social media. All students are automatically enrolled in Warrior Shield through their myCT email account.

Connect to Warrior Shield by [911Cellular](https://portal.publicsafetycloud.net/Account/Login) [<https://portal.publicsafetycloud.net/Account/Login>] to change where you receive your alerts or to opt out. By staying enrolled in Warrior Shield, university officials can quickly pass on safety-related information, regardless of your location.

COURSE INFORMATION

Course Overview and description

The course covers fundamental concepts and principles of Data Mining and its role in supporting/enhancing organizational-decision making and predictions. Topics covered include Business Analytics, Data Science, Big Data, Data Visualization, Dimension Reduction, Evaluating Predictive Performance, Multiple Linear Regression, Classification with k-Nearest Neighbors algorithm and Naïve Bayes Classifier, Logistic Regression, Association Rules and Cluster Analysis.

Course Objective and Student Learning Outcomes

Successful completion of this course should enable student to:

- Describe Data Science, Business Analytics and Big Data with their role in the corporate world.
- Describe the data mining process.
- Visualize data and explain insights via data visualization.
- Describe curse of dimensionality.
- Evaluate predictive performance.
- Describe, implement and explain the results of multiple linear regression.
- Describe and solve classification problems via Naïve Bayes and Discriminant analysis.
- Describe, implement and explain the results of logistic regression.
- Describe neural nets.
- Describe association rules and implement association rule mining.
- Describe and implement cluster analysis.

Competency Goals Statements (certification or standards)

None

Required Reading and Textbook(s)

Data Mining for Business Analytics – Concepts, techniques, and applications in R, Wiley, ISBN: 9781118879368

Optional (Reference) Textbook:

Introduction to Data Mining, Second Edition, Pearson, ISBN: 9780133128901

R for Data Science, O'Reilly, ISBN: 9781491910399

Note: A student of this institution is not under any obligation to purchase a textbook from a university-affiliated bookstore. The same textbook may also be available from an independent retailer, including an online retailer.

Reading Assignments: All assigned chapters along with assigned readings will be used for class and/or canvas discussions. Students are expected to study the assigned readings before each class session.

Supplementary Material: The course textbook will be supplemented with other materials that may include research papers readings, handouts, oral presentations, industry articles, videos including TED talks, research paper readings, case studies, power point presentations etc. Students must know how to search for and download research papers / articles from the Web (more specifically from Google scholar (www.scholar.google.com))

COURSE REQUIREMENTS

Examinations: There will be three exams. Each exam will be worth 20 points. The exams will have two parts: multiple choice questions (answered online via Canvas) and programming problems (executable code submitted via Canvas).

The exams will all be open book / open notes and will be available via Canvas. You may also use your submitted home works but use of Internet not allowed. For the exams, you will be required to know not only the material from the textbook chapters, but also material from the class lectures such as power point slides and any supplementary/additional material provided as well. You must be physically present in-class to take all the exams (taking exams from home is not permitted under any circumstances). You must also know how to write and debug R programming code independently. If you cannot take the in-class exam, inform me via email to plan to take on-paper scantron based exams within 10 days of in-class exam(s), during my office hours.

Individual home works: There will be four individual home works. Students will get approximately two weeks to work on the individual homework assignments. Each homework will be worth 5 points and may cover one or more of the following topics (in no order):

- Basic R programming
- Data visualization
- Multiple linear and logistic regression
- Classification via Naïve Bayes and discriminant analysis
- Discriminant analysis
- Neural nets
- Association rules and cluster analysis

Home works turned in after due date are considered late. **1 point deducted for each day the home work is late.** Special circumstances need to be discussed with the instructor ahead of time when possible.

Semester wide team research project: There will be one semester long data mining project worth 20 points (15 points for the project report / research paper and 5 points for project proposal). Students can self-select into teams of 2 to 3 members using the Canvas team tool. Students must choose their teammates before the end of the third week. In case of students not able to form groups, instructor will assign students to random groups during the fourth week. The team must choose a freely available secondary data source from the Web (more on this during the second or third week of the class) and pose some research questions. There are many sources of secondary data available on the Web, you are free to explore and are not restricted to a specific type of data or application domain. Students are required to form a team and submit project proposal by the end of fourth week (before the mid-term exam).

The aim of the project is to use the data mining techniques learned during the class to test the verifiability of the proposed research question(s). Students are not required to turn in the data or the programming code but must submit a written report of their findings in the form of a short research paper (guidelines for the team project write up will be provided soon). Although it is not a requirement to come with new research problem(s), students are encouraged to do so. **1 point will be deducted for each day the team project proposal is late.**

Projects turned in after due date are considered late. **5 points deducted for each day the project is late.** Special circumstances need to be discussed with the instructor ahead of time when possible. Each team may select a leader, and the team will have meetings outside class time either online or face to face. I expect each team member to fully participate on team projects. All team members will receive the same grade for the team projects. However, I reserve the right to make exceptions to that practice as circumstances such as performance imbalance or communication issues warrant. Poor individual contributions to the team project as noted by your team members will result in a poorer individual grade on the team project.

Grading Criteria Rubric and Conversion

Item	Quantity	Points	Total / Percent
Exam(s)	3	15	45
Home works	4	10	40
Semester long team project proposal	1	5	5
Semester long team project	1	10	10
Total			100

Exams, quizzes, assignments and individual term project will receive a numeric score (0-100) each. These scores will be converted to points and totaled to ultimately be converted to letter grade of A, B, C, D, or F as shown in the example below:

	Percent earned by Student	Max points	Points to be added to the final grade
Exam 1	78%	15	11.7
Exam 2	85%	15	12.75
Exam 3	90%	15	13.5
Homework 1	100%	10	10
Homework 2	75%	10	7.5
Homework 3	80%	10	8
Homework 4	90%	10	9
Semester long team project proposal	100%	5	5
Semester long team project report	80%	15	12
TOTAL		100	88.45

89.5 – 100 = A

79.5 – 89.49999 = B

69.5 – 79.49999 = C

59.5 – 69.49999 = D

Below 59.5 = F

Grades will not be curved (in the above example, the student with 88.45 will receive a B NOT an A).

Posting of Grades

All student grades will be posted on the Canvas Grade Book and students should monitor their grading status through this tool. Grades will be posted within 10 days after the due date.

COURSE OUTLINE AND CALENDAR

Complete Course Calendar (subjected to change)

Date	Content	Readings	Assignments
August 26, 28	Course Welcome, Introduction, and overview of course objectives, and expectations. Understanding Data Science, Data Analytics and Big Data.	Read Syllabus in-depth. Chapter 1 – Introduction	

September 2, 4	Overview of data mining process	Chapter 2	Homework 1 on R programming available
September 9, 11	Data visualization	Chapter 3	September 11: Deadline to drop 16-week classes with no record.
September 16, 18	Dimension reduction	Chapter 4	Homework 1 due on September 18
September 23, 25	Evaluating predictive performance	Chapter 5	
September 30	Exam – 1	Chapter 1 – 5	
October 2, 7	Multiple linear regression	Chapter 6	
October 9, 14	Logistic Regression	Chapter 10	Homework 2 on linear and logistic regression available
October 16, 21	k-Nearest Neighbors.	Chapter 7	
October 23, 28	Naïve Bayes Classifier.	Chapter 8	Homework 2 due on October 28.
October 30, November 4	Classification Trees	Chapter 9	
November 6	Exam – 2	Chapter 6 – 10	November 8: Deadline to drop 16-week classes with a quit(Q) or withdraw(W).
November 11, 13	Neural Nets	Chapter 11	Homework 3 on neural nets and/or discriminant analysis available.
November 18, 20	Discriminant analysis	Chapter 12	
November 25, 27	Cluster analysis	Chapter 15	Homework 3 due on November 27. Homework 4 on cluster analysis available.

November 28, 29: No class (Thanksgiving – University closed).			
December 2, 4	Association rules and collaborative filtering	Chapter 14	Homework 4 due December 4.
December 9	Exam – 3	Chapters – 11, 12, 14 and 15	Team Term Project Due on December 10 mid night.

Important University Dates

Refer to University official academic calendar at <https://www.tamuct.edu/registrar/academic-calendar.html>

TECHNOLOGY REQUIREMENTS AND SUPPORT

Technology Requirements

Latest version of R programming language and RStudio integrated development environment needs to be downloaded and installed for successful completion of home works, team project and exams (if you will be taking exams on your laptop). Both R and RStudio are opensource freely available from <https://www.r-project.org/> and <https://www.rstudio.com/products/rstudio/download/> respectively. The University’s computer labs are equipped with these software’s.

This course will use the A&M-Central Texas Instructure Canvas learning management system. Logon to A&M-Central Texas Canvas [<https://tamuct.instructure.com/>] or access Canvas through the TAMUCT Online link in myCT [<https://tamuct.onecampus.com/>]. You will log in through our Microsoft portal.

Username: Your MyCT email address. Password: Your MyCT password

Canvas Support

Use the Canvas Help link, located at the bottom of the left-hand menu, for issues with Canvas. You can select “Chat with Canvas Support,” submit a support request through “Report a Problem,” or call the Canvas support line: 1-844-757-0953.

For issues related to course content and requirements, contact your instructor.

Other Technology Support

For log-in problems, students should contact Help Desk Central 24 hours a day, 7 days a week

Email: helpdesk@tamu.edu

Phone: (254) 519-5466

Web Chat: [<http://hdc.tamu.edu>]

Please let the support technician know you are an A&M-Central Texas student.

UNIVERSITY RESOURCES, PROCEDURES, AND GUIDELINES

Drop Policy

If you discover that you need to drop this class, you must complete a [Drop Request Form](https://www.tamuct.edu/registrar/docs/Drop_Request_Form.pdf) [https://www.tamuct.edu/registrar/docs/Drop_Request_Form.pdf].

Professors cannot drop students; this is always the responsibility of the student. The Registrar's Office will provide a deadline on the Academic Calendar for which the form must be completed, signed and returned. Once you return the signed form to the Registrar's Office, you must go into Warrior Web and confirm that you are no longer enrolled. If you still show as enrolled, FOLLOW-UP with the Registrar's Office immediately. You are to attend class until the procedure is complete to avoid penalty for absence. Should you miss the drop deadline or fail to follow the procedure, you will receive an F in the course, which may affect your financial aid and/or VA educational benefits.

Academic Integrity

Texas A&M University -Central Texas values the integrity of the academic enterprise and strives for the highest standards of academic conduct. A&M-Central Texas expects its students, faculty, and staff to support the adherence to high standards of personal and scholarly conduct to preserve the honor and integrity of the creative community. Academic integrity is defined as a commitment to honesty, trust, fairness, respect, and responsibility. Any deviation by students from this expectation may result in a failing grade for the assignment and potentially a failing grade for the course. Academic misconduct is any act that improperly affects a true and honest evaluation of a student's academic performance and includes, but is not limited to, cheating on an examination or other academic work, plagiarism and improper citation of sources, using another student's work, collusion, and the abuse of resource materials. All academic misconduct concerns will be reported to the university's Office of Student Conduct. Ignorance of the university's standards and expectations is never an excuse to act with a lack of integrity. When in doubt on collaboration, citation, or any issue, please contact your instructor before taking a course of action.

For more [information regarding the Student Conduct process](https://www.tamuct.edu/student-affairs/student-conduct.html), [<https://www.tamuct.edu/student-affairs/student-conduct.html>].

If you know of potential honor violations by other students, you may [submit a report](https://cm.maxient.com/reportingform.php?TAMUCentralTexas&layout_id=0), [https://cm.maxient.com/reportingform.php?TAMUCentralTexas&layout_id=0].

Academic Accommodations

At Texas A&M University-Central Texas, we value an inclusive learning environment where every student has an equal chance to succeed and has the right to a barrier-free education. The Office of Access and Inclusion is responsible for ensuring that students with a disability receive equal access to the university's programs, services and activities. If you believe, you have a disability requiring reasonable accommodations please contact the Office of Access and Inclusion, WH-212; or call (254) 501-5836. Any information you provide is private and confidential and will be treated as such.

For more information please visit our [Access & Inclusion](#) Canvas page (login required)

[<https://tamuct.instructure.com/courses/717>]

Important information for Pregnant and/or Parenting Students

Texas A&M University-Central Texas supports students who are pregnant and/or parenting. In accordance with requirements of Title IX and related guidance from US Department of Education's Office of Civil Rights, the Dean of Student Affairs' Office can assist students who are pregnant and/or parenting in seeking accommodations related to pregnancy and/or parenting. Students should seek out assistance as early in the pregnancy as possible. For more information, please visit [Student Affairs](https://www.tamuct.edu/student-affairs/index.html) [<https://www.tamuct.edu/student-affairs/index.html>]. Students may also contact the institution's Title IX Coordinator. If you would like to read more about these requirements and guidelines online, please visit the website [<http://www2.ed.gov/about/offices/list/ocr/docs/pregnancy.pdf>].

Title IX of the Education Amendments Act of 1972 prohibits discrimination on the basis of sex and gender—including pregnancy, parenting, and all related conditions. A&M-Central Texas is able to provide flexible and individualized reasonable accommodation to pregnant and parenting students. All pregnant and parenting students should contact the Associate Dean in the Division of Student Affairs at (254) 501-5909 to seek out assistance. Students may also contact the University's Title IX Coordinator.

Tutoring

Tutoring is available to all A&M-Central Texas students, both on-campus and online. Subjects tutored on campus include Accounting, Advanced Math, Biology, Finance, Statistics, Mathematics, and Study Skills. Tutors are available at the Tutoring Center in Warrior Hall, Suite 111. Tutor.com tutoring **will not offer writing support** beginning August 1, 2019, but will continue to offer other tutoring support.

If you have a question regarding tutor schedules, need to schedule a tutoring session, are interested in becoming a tutor, or have any other question, contact Academic Support Programs at (254) 519-5796, or by emailing Dr. DeEadra Albert-Green at deeadra.albertgreen@tamuct.edu.

Chat live with a tutor 24/7 for almost any subject from on your computer! Tutor.com is an online tutoring platform that enables A&M-Central Texas students to log in and receive online tutoring support at no additional cost. This tool provides tutoring in over 40 subject areas. Access Tutor.com through Canvas.

University Writing Center

Located in Warrior Hall 416, the University Writing Center (UWC) at Texas A&M University–Central Texas (TAMUCT) is a free workspace open to all TAMUCT students from 10:00 a.m.-5:00 p.m. Monday thru Thursday with satellite hours in the University Library Monday thru Thursday from 6:00-9:00 p.m. This semester, the UWC is also offering online only hours from 12:00-3:00 p.m. on Saturdays.

Tutors are prepared to help writers of all levels and abilities at any stage of the writing process. While tutors will not write, edit, or grade papers, they will assist students in developing more

effective composing practices. By providing a practice audience for students' ideas and writing, our tutors highlight the ways in which they read and interpret students' texts, offering guidance and support throughout the various stages of the writing process. In addition, students may work independently in the UWC by checking out a laptop that runs the Microsoft Office suite and connects to WIFI, or by consulting our resources on writing, including all of the relevant style guides. Whether you need help brainstorming ideas, organizing an essay, proofreading, understanding proper citation practices, or just want a quiet place to work, the UWC is here to help!

Students may arrange a one-to-one session with a trained and experienced writing tutor by visiting the UWC during normal operating hours (both half-hour and hour sessions are available) or by making an appointment via [WCOonline](https://tamuct.mywconline.com/) [https://tamuct.mywconline.com/]. In addition, you can email Dr. Bruce Bowles Jr. at bruce.bowles@tamuct.edu if you have any questions about the UWC and/or need any assistance with scheduling.

University Library

The University Library provides many services in support of research across campus and at a distance. We offer over 200 electronic databases containing approximately 250,000 eBooks and 82,000 journals, in addition to the 85,000 items in our print collection, which can be mailed to students who live more than 50 miles from campus. Research guides for each subject taught at A&M-Central Texas are available through our website to help students navigate these resources. On campus, the library offers technology including cameras, laptops, microphones, webcams, and digital sound recorders.

Research assistance from a librarian is also available 24 hours a day through our online chat service, and at the reference desk when the library is open. Research sessions can be scheduled for more comprehensive assistance, and may take place on Skype or in-person at the library. Assistance may cover many topics, including how to find articles in peer-reviewed journals, how to cite resources, and how to piece together research for written assignments.

Our 27,000-square-foot facility on the A&M-Central Texas main campus includes student lounges, private study rooms, group work spaces, computer labs, family areas suitable for all ages, and many other features. Services such as interlibrary loan, TexShare, binding, and laminating are available. The library frequently offers workshops, tours, readings, and other events. For more information, please visit our [Library website](http://tamuct.libguides.com/index) [http://tamuct.libguides.com/index].

OPTIONAL POLICY STATEMENTS

A Note about Sexual Violence at A&M-Central Texas

Sexual violence is a serious safety, social justice, and public health issue. The university offers support for anyone struggling with these issues. University faculty are mandated reporters, so if someone discloses that they were sexually assaulted (or a victim of Domestic/Dating Violence or Stalking) while a student at TAMUCT, faculty members are required to inform the Title IX Office. If you want to discuss any of these issues confidentially, you can do so through Student Counseling (254-501-5955) located on the second floor of Warrior Hall (207L).

Sexual violence can occur on our campus because predators often feel emboldened, and victims often feel silenced or shamed. It is incumbent on ALL of us to find ways to actively create environments that tell predators we don't agree with their behaviors and tell survivors we will support them. Your actions matter. Don't be a bystander; be an agent of change. For additional information on campus policy and resources visit the [Title IX webpage](https://www.tamuct.edu/departments/compliance/titleix.php) [<https://www.tamuct.edu/departments/compliance/titleix.php>].

Behavioral Intervention

Texas A&M University-Central Texas cares about the safety, health, and well-being of its students, faculty, staff, and community. If you are aware of individuals for whom you have a concern, who are exhibiting behaviors that pose a threat to safety, or individuals causing a significant disruption to our community, please make a referral to the Behavioral Intervention Team. You can complete the [referral](https://cm.maxient.com/reportingform.php?TAMUCentralTexas&layout_id=2) online [https://cm.maxient.com/reportingform.php?TAMUCentralTexas&layout_id=2].

Anonymous referrals are accepted. Please see the [Behavioral Intervention Team](https://www.tamuct.edu/student-affairs/bat.html) website for more information [<https://www.tamuct.edu/student-affairs/bat.html>]. If a person's behavior poses an imminent threat to you or another, contact 911 or A&M-Central Texas University Police at 254-501-5800.

INSTRUCTOR POLICIES

Policies related to absence, grading, etc.

- You are responsible for all class material presented during an absence.
- Home works must be submitted through Canvas and due on the mid night (11:59 PM) of specified due date/time. No email submissions will be accepted under any circumstances.
- Late penalties will be applied to items submitted after due dates as per the per day late submit penalty guidelines stated above.
- If you cannot take in-class exam, you must inform me at least one week beforehand and must plan to take the exam during my office hours within 10 days of exam due dates.

My personal statement

- You will receive feedback in the form of graded assignments within 10 days after the due date.
- I want you to read the feedback that I provide to you (your personal grading notes and Canvas emails).
- I am almost always available via email and typically respond within 24 hours except on weekends.
- I prefer email to phone conversations.

- I reserve the right to modify the course syllabus during the semester for the benefit of the students.
- I reserve the right to supplement materials presented in the text with additional course material that may help the students to understand the topic better.
- I reserve the right to modify grading policy rubrics. Any change to grading rubrics will be applied to current and possible future assignments.

Copyright Notice

Students should assume that all course material is copyrighted by the respective author(s). Reproduction of course material is prohibited without consent by the author and/or course instructor. Violation of copyright is against the law and Texas A&M University-Central Texas' Code of Academic Honesty. All alleged violations will be reported to the Office of Student Conduct.