# CIS 4303 – 110, Data Mining

## Fall 2020

Texas A&M University-Central Texas

## COURSE DATES, MODALITY, AND LOCATION

Course meeting dates: August 24, 2020 – Dec 11, 2020
Class meeting building and room number: Founders Hall 207
Class meeting day and time: Mon 2:00 PM to 3:15 PM, Wed 2:00 PM to 3:15 PM virtually via a recorded lecture (I will be available via Canvas and University email to answer any questions)

## INSTRUCTOR AND CONTACT INFORMATION

**Instructor: Rahul Dwivedi**
**Office: Founders Hall 323H**
**Phone: NA**
**Email:** rahul.dwivedi@tamuct.edu (preferred) or Canvas inbox.

## Office Hours

*Since the course is 50% online, concerns of students will be addressed via email. Students can send emails with their concerns or questions related with any aspect of the course any time between 9 AM – 5 PM Monday to Thursday.  If there's something which needs to be addressed via virtual face-to-face meeting, students can request WebEx (*https://tamuct.onecampus.com/task/all/webex*) meeting(s) as well.*

## Student-instructor interaction

*I check my University email many times a day and typically responds within 24 hours of receiving emails on weekdays. Email is also the best mode of communication.*

## Mode of instruction and course access

This a 50% online 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, video lectures, learning modules, homework assignments, exams and other course related documents will be posted on Canvas throughout the week. Each student is responsible for the posted material and should check Canvas several times a week for updates.

## 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] 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.

## COVID-19 SAFETY MEASURES

To promote public safety and protect students, faculty, and staff during the coronavirus pandemic, Texas A&M University-Central Texas has adopted policies and practices to minimize virus transmission. All members of the university community are expected to adhere to these measures to ensure their own safety and the safety of others. Students must observe the following practices while participating in face-to-face courses, course-related activities (office hours, help sessions, transitioning to and between classes, study spaces, academic services, etc.) and co-curricular programs:

- Self-monitoring—Students should follow CDC recommendations for self-monitoring. Students who have a fever or exhibit symptoms of COVID-19 should participate in class remotely and should not participate in face-to-face instruction. Students required to quarantine must participate in courses and course-related activities remotely and must not attend face-to-face course activities. Students should notify their instructors of the quarantine requirement. Students under quarantine are expected to participate in courses and complete graded work unless they have symptoms that are too severe to participate in course activities.

- Face Coverings— Face coverings must be worn inside of buildings and within 50 feet of building entrances on the A&M-Central Texas Campus. This includes lobbies, restrooms, hallways, elevators, classrooms, laboratories, conference rooms, break rooms, non-private office spaces, and other shared spaces. Face coverings are also required in outdoor spaces where physical distancing is not maintained. The university will evaluate exceptions to this requirement on a case by case basis. Students can request an exception through the Office of Access and Inclusion in Student Affairs.

  o If a student refuses to wear a face covering, the instructor should ask the student to leave and join the class remotely. If the student does not leave the class, the faculty member should report that student to the Office of Student Conduct. Additionally, the faculty member may choose to teach that day's class remotely for all students.

- Physical Distancing—Physical distancing must be maintained between students, instructors, and others in the course and course-related activities.

- Classroom Ingress/Egress—Students must follow marked pathways for entering and exiting classrooms and other teaching spaces. Leave classrooms promptly after course activities have concluded. Do not congregate in hallways and maintain 6-foot physical distancing when waiting to enter classrooms and other instructional spaces.

- The university will notify students in the event that the COVID-19 situation necessitates changes to the course schedule or modality.

**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, Multiple Linear Regression, Classification with k-Nearest Neighbors algorithm and Naïve Bayes Classifier, Logistic Regression, Neural Networks, 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.
- Describe, implement and explain the results of multiple linear regression.
- Describe and solve classification problems via Naïve Bayes.
- 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:**

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 homework's but use of Internet is 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.

**Individual homework's:** There will be four individual homework's. 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
- Neural nets
- Association rules and cluster analysis

Homework's turned in after due date are considered late. **1 point will be deducted for each day the homework is late.** Special circumstances need to be discussed with the instructor ahead of time when possible.

**Semester wide individual research project:** There will be one semester long data mining project worth 20 points. Student 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.

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.

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.

**Grading Criteria Rubric and Conversion**

| Item | Quantity | Points | Total / Percent |
|------|----------|--------|-----------------|
| Exam(s) | 3 | 20 | 60 |
| Homework's | 4 | 5 | 20 |
| Semester long team project | 1 | 20 | 20 |
| 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% | 20 | 15.6 |
| Exam 2 | 85% | 20 | 17 |
| Exam 3 | 90% | 20 | 18 |
| Homework 1 | 100% | 5 | 10 |
| Homework 2 | 75% | 5 | 3.75 |
| Homework 3 | 80% | 5 | 4 |
| Homework 4 | 90% | 5 | 9 |
| Semester long team project report | 60% | 20 | 12 |
| TOTAL | | 100 | 89.35 |

**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 89.35 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**

| Date | Content | Readings | Assignments |
|---|---|---|---|
| August 24, 26 | 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 | |
| August 31, September 2 | Overview of data mining process | Chapter 2 | Homework 1 on R programming available |
| September 7, 9 | Data visualization | Chapter 3 | September 9: Deadline to drop 16-week classes with no record. |
| September 14, 16 | Dimension reduction | Chapter 4 | Homework 1 due on September 14 |
| September 21 | Exam # 1 | Chapter 1 – 4 | |
| September 28 | Multiple linear regression | Chapter 6 | Homework 2 on linear and logistic regression available. |
| October 5, 7 | Logistic regression | Chapter 10 | |
| October 12, 14 | k-Nearest Neighbors. | Chapter 7 | Homework 2 due on October 12. Homework 3 based on k-NN and Naïve Bayes classifiers |

| | | | available on October 12. |
|---|---|---|---|
| October 19, 21 | Naïve Bayes Classifier. | Chapter 8 | |
| October 26, 28 | Classification Trees | Chapter 9 | Homework 3 due on October 26. |
| November 2 | Exam # 2 | Chapter 6 – 10 | November 6: Deadline to drop 16-week classes with a quit(Q) or withdraw(W). |
| November 4, 9 | Neural Nets | Chapter 11 | |
| November 11, 16 | Cluster analysis | Chapter 15 | Homework 4 on neural nets and cluster analysis available on November 16. |
| November 18, 23 | Association rules | Chapter 14 | |
| November 30, December 2 | Reserve Week | | Homework 4 due November 30. |
| December 7, 2020 | Exam # 3 | Chapters – 11, 14 and 15 | Individual Project Due on December 8th mid night. |

**Important University Dates**
*Refer to: 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 homework's, individual project and exams (if you are bringing in your laptop for class work and taking exams). 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. We strongly recommend the latest versions of Chrome or Firefox browsers. Canvas no longer supports any version of Internet Explorer.

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.

**Online Proctored Testing**

A&M-Central Texas uses Proctorio for online identity verification and proctored testing. This service is provided at no direct cost to students. If the course requires identity verification or proctored testing, the technology requirements are: Any computer meeting the minimum computing requirements, plus web camera, speaker, and microphone (or headset). Proctorio also requires the Chrome web browser with their custom plug in.

**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 the Drop Request Dynamic Form through Warrior Web.

[https://dynamicforms.ngwebsolutions.com/casAuthentication.ashx?InstID=eaed95b9-f2be-45f3-a37d-46928168bc10&targetUrl=https%3A%2F%2Fdynamicforms.ngwebsolutions.com%2FSubmit%2FForm%2FStart%2F53b8369e-0502-4f36-be43-f02a4202f612].

Faculty 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. Once you submit the completed 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, working with others in an unauthorized manner, 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 referred 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](https://tamuct.instructure.com/courses/717) Canvas page (log-in 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, on a remote online basis. Visit the Academic Support Community in Canvas to view schedules and contact information. Subjects tutored on campus include Accounting, Advanced Math, Biology, Finance, Statistics, Mathematics, and Study Skills. Tutors will return at the Tutoring Center in Warrior Hall, Suite 111 in the Fall 2020. Student success coaching is available online upon request.

If you have a question regarding tutor schedules, need to schedule a tutoring session, are interested in becoming a tutor, success coaching, or have any other question, contact Academic Support Programs at (254) 501-5836, visit the Office of Student Success at 212F Warrior Hall, or by emailing studentsuccess@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 except writing support. Access Tutor.com through Canvas.

**University Writing Center**

The University Writing Center (UWC) at Texas A&M University–Central Texas (TAMUCT) is a free service open to all TAMUCT students. For the Fall 2020 semester, all services will be online as a result of the COVID-19 pandemic. The hours of operation are from 10:00 a.m.-5:00 p.m. Monday thru Thursday with satellite hours online Monday thru Thursday from 6:00-9:00 p.m. The UWC is also offering 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. 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. While tutors will not write, edit, or grade papers, they will assist students in developing more effective composing practices. 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

making an appointment via WCOnline [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].

**For Fall 2020, all reference service will be conducted virtually. Please go to our Library website [http://tamuct.libguides.com/index] to access our virtual reference help and our current hours.**

---

**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/compliance/titleix.html) [https://www.tamuct.edu/compliance/titleix.html].

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, please make a referral to the Behavioral Intervention Team. Referring your concern shows you care. 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.

## OTHER POLICIES

## INSTRUCTOR POLICIES

### Policies related to absence, grading, etc.

- ☐ You are responsible for all class material presented during an absence.
- ☐ Homework's 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.

### My personal statement

- ☐ You will receive feedback in the form of graded assignments within 2 weeks 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.