MATH 5364-010 Data Science 1 - Syllabus Fall 2020

Instructor: Dr. Scott Cook	**E-mail: scook@tarleton.edu**
Class Time: 9:25-10:50 TR	**Classroom: Math Bldg 134 (see F2F restrictions)**
Phone: 254-968-1958	**Office Hours: M,W 11-1 & by appt**
Math Dept Phone: 254-968-9168	**Schedule: https://calendly.com/scook-10/15min**
Zoom for Class Meetings **https://tarleton.zoom.us/j/94853849160** **password = math**	**Zoom for Office Hours** **https://tarleton.zoom.us/my/scook** **password = math**

Course Description

This course centers on the identification, exploration, and description of new patterns contained within data sets using appropriate software. Selected topics will be chosen from data exploration, classification, cluster analysis, and model evaluation and comparison.

Texts, Materials, or Equipment

- GitHub account
- Google Drive account
- Webcam & stable internet
- Canvas - https://www.tarleton.edu/oiss/canvas-login.html
- We will use a lot of free, open-source resources (and likely no paid ones)

Topics

The field of data science is ENORMOUS. In essence, any time we try to detect patterns in data, we are doing data science. Thus, it is laughable to try to cover it all in 2 courses. So, we will spend the first semester carefully building a strong foundation of core data science concepts and tools and the second semester rapidly surveying as many of the more specialized and cutting edge tools and algorithms as we can.

Core topics for the first semester include, but are not limited to:

- Packages - Python, Numpy, Pandas, Scikit Learn, Jupyter
- Visualization & Presentation - Matplotlib, Seaborn, Bokeh, Holoviews, interactive widgets
- Platforms - Google Colab, GitHub
- Supervised Learning: linear regression, *k*-nearest neighbors, naive Bayes, support vector machines, Bayesian models, tree-based algorithms (decision trees, random forests, XGboost, Light GBM, Catboost), artificial neural networks (deep learning), ensemble methods
- Unsupervised Learning - dimensionality reduction (principal components, multidimensional scaling, locally linear embedding, UMAP, etc) and clustering (k-means, gaussian mixtures, hierarchical clustering, DBSCAN)
- Concepts - overfitting vs underfitting, curse of dimensionality, cross-validation, eager vs lazy, class imbalance problems, data cleaning and preparation, feature extraction and selection
- Model tuning & evaluation - confusion matrices, accuracy/precision/recall/F_1, ROC & PR curves, cost sensitive learning, hyper-parameter optimization

As time allows, I also want to touch on recent evolutions of these tools, such as xarray, jax, dask, word embeddings, MCMC techniques, etc.

Expectations, Mindset, and Time Management

In most courses, instructors carefully linearize topics so that you've covered everything needed before starting an assignment. But real, interesting projects worth doing are almost never like this, especially in a field as broad, dynamic, and niche-filled as data science. The pace of expansion, evolution, and specialization means your data science projects will constantly demand that you teach yourself new stuff for which there is no book or course.

Thus, I deliberately leave this course "loose". We're using a collection of internet resources instead of following a textbook. Your project and homework will often require you to learn stuff we haven't covered in class through online tutorials and good-ole trial-and-error. Please collaborate with each other and me. Adjust your mindset now to expect "non-linearity" in the course.

Homework and projects may take a long time, especially if you hit code bugs. It is never smart to procrastinate and try to cram everything in just before it is due. But that is magnified in this course because there are so many bugs to squash and each could take days. Start on everything as soon as it is posted and plan to work over many sessions to maximize bug squashing time.

Assignments

Homework will be assigned approximately weekly and due 1-2 weeks later. Occasional short quizzes may be used to reinforce terminology and core concepts. Assignments will be distributed and collected via GitHub. (I'm experimenting with GitHub Classroom this semester.)

Collaboration is extremely beneficial, especially when writing coding. The mathematical ideas are complex; writing them into working code is even more complex; dealing with flaws in your data and bugs in your code even more. You're going to get stuck and frustrated. When that happens, most of us get "tunnel vision" that keeps us from seeing our own errors. Work with your classmates to help each get out of the tunnel.

Thank your collaborators in your final submission; you won't lose credit. Giving credit to others when it's due shows you are a professional that can work effectively in a team.

I strongly urge you to work in the graduate office and computer lab as much as possible to facilitate collaboration. Obviously, that's more delicate in Fall 2020 than before, but try to find a safe balance.

Project

You will do a capstone project using techniques from the course, due at the end of the semester. I will post detailed instructions shortly.

Start looking for your datasets and project ideas **<u>now</u>**! You will be stunned how long it takes to clean and prepare the data and hone your questions to a "Goldilocks set" - not too simple, not too complex, just right.

Start now! Do not wait because the end of the semester will be stressful and demanding. We will cover necessary topics throughout the semester, so you probably can't finish before we cover them. But you can make a lot of progress. I may ask for periodic status reports to help refine your ideas, avoid pitfalls, solve technical challenges, and keep you on track.

Exams
There will be a comprehensive final exam Thursday, December 3, 8:00-10:00. This is intended to provide a reason to consolidate and clarify the core concepts and terms from the course as preparation for future job interviews and your comprehensive oral exams.

I prefer to hold this in-person if logistics allow, but we will explore options as the time approaches.

***Grading Policy*:** The guaranteed grade weights and cutoffs are listed below. At my sole discretion, I may curve the course by relaxing the cutoffs at the end of the semester.

Homework/Quiz: 50%, Project: 40%, Final Exam: 10%
F: [0%, 60%), D: [60%, 70%), C: [70%, 80%), B: [80%, 90%), A: [90%, ∞)

Covid-19 Policies - My Summary

- **MASKS & SOCIAL DISTANCING ARE REQUIRED** for all face-to-face (F2F) contact throughout the Fall 2020 semester
- All students must complete Covid-19 training before the first day of class or returning to campus: https://www.tarleton.edu/oiss/canvas-login.html
- Classes end 5 minutes early
- Zoom (no F2F) office hours only; faculty offices are now 1 occupant only; schedule at: **https://calendly.com/scook-10/15min**
- Covid forced us to implement a lot of new things at the same time.
    - Engineers, scientists, and designers know this is not ideal because it greatly increases the risk of things going wrong. But, we don't have any other choice.
    - Please extend grace if things don't work perfectly.
    - We'll try to fix it as fast as we can with as little impact on you as possible.
    - This will be a hard semester.
    - Please don't make it harder by resisting the safety measures (masks, distancing, etc).
    - Even if you are not concerned or convinced about Covid, people around you are.
    - Your choices/rights can harm (and kill) other people.

HyFlex in Math 5364
Due to Covid-19, Tarleton is following the HyFlex model to provide flexibility in this time of uncertainty. Though this course has online elements, it is NOT an online course. You may flex among the following options as necessary:

- Synchronous = at scheduled class time
    - F2F = face-to-face in the assigned classroom - see seating chart in Canvas
    - Zoom
- Asynchronous = not at scheduled class time
    - Recordings of class meetings (not pre-recorded lectures as in fully online courses)
    - Online resources (Canvas, Connect, etc)
- You should always attend synchronously (F2F or Zoom) unless you have an unavoidable conflict.
    - You will achieve your best learning and grade outcomes through asking and learning from questions and discussion in class.
    - I know schedules are unpredictable and unavoidable conflicts arise
    - I will work hard to make the asynchronous experience as equivalent as possible to synchronous. I want to help you keep up and remain successful.
    - Think of asynchronous as a backup option because you can't ask live questions.
    - Please email & schedule office hours with any questions or concerns.

University Policies

Safety Measures

Tarleton State University has adopted policies and practices for the Fall 2020 term to limit transmission of the novel coronavirus. Students are required to observe the following practices while participating in face-to-face courses and course-related activities (office hours, moving between classes, study spaces, academic services, etc.):

- Self-monitoring – Students should follow CDC recommendations for self-monitoring. **Students who exhibit symptoms of COVID-19 (with or without fever) should participate in class remotely and should not participate in face-to-face instruction.** See https://www.tarleton.edu/roadmap/personal-responsibilities/ for more information. Students who test positive for COVID-19 or experience symptoms consistent with COVID-19 are required to self-report to Tarleton State University via this form.
- Face coverings – All students must properly wear face coverings in all public spaces on campus, including classrooms. If a student refuses to wear a face covering, the instructor will ask the student to leave and join the class remotely. Any student refusing to comply will be reported to the Dean of Students Administrative Office via the Student Affairs Incident Reporting Form. 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 course and course-related activities.
- Classroom Entrance and Exit – Students should leave classrooms promptly after class activities have concluded each day. Students should not congregate in hallways or other areas and should maintain a safe physical distance when waiting to enter classrooms and other instructional areas.

Personal Illness and Quarantine/Isolation

Students who are required to quarantine (see https://www.tarleton.edu/roadmap/isolation-v-quarantine/) must participate in course and course-related activities remotely and **must not attend face-to-face course activities.** Students in quarantine are expected to participate in courses and course activities/assignments unless they have symptoms too severe to participate. Students placed in isolation should contact the instructor about individual participation in relation to severity of illness. Students who test positive for COVID-19 or who are experiencing symptoms consistent with COVID-19 are required to self-report to the Dean of Students Administrative Office through the COVID-19 Report Form. For any questions or concerns, please contact the Dean of Students Administrative Office at 254-968-9080.

Blended Hybrid-HyFlex Course Delivery

Blended Hybrid-HyFlex courses are designed so that students can choose to attend courses face-to-face (with the potential for rotation to maintain a safe physical distance), at the same time as the face-to-face class meetings but from a different location (remote synchronous), or remotely at a later time by viewing the recorded course meeting (remote asynchronous). All courses will appear in the Canvas Learning Management System (LMS) to maximize access to course materials and other important course related activities. Students can choose to attend via any of the three modalities at any time.

Note, however, that programs governed by licensure and/or accreditation/certification requirements may require students to attend face-to-face laboratories, simulations, and clinical experiences to progress through the program and successfully graduate with eligibility for the licensure or certification examination.

To make course meetings accessible asynchronously, class meetings will be recorded and shared. The class recordings will be shared with students only in the individual section in which it was recorded to avoid violations of the Family Educational Rights and Privacy Act (FERPA).

Study Aids

- The services below are geared more for lower level courses, but offer some limited support for Math 3310.
- The Mathematics Clinic offers free tutoring to students enrolled in mathematics courses. See https://www.tarleton.edu/math/math_clinic.html (time & modality may vary due to staffing and capacity restrictions).
- The university offers several programs through which students may obtain free or reduced-fee private tutoring. Interested students should visit the Academic Resource Center https://www.tarleton.edu/tlc/tutoring/overview.html for more details.

Student Success Statement - ADA, Services for Students with Disabilities: It is the policy of Tarleton State University to comply with the Americans with Disabilities Act (www.ada.gov) and other applicable laws. If you are a student with a disability seeking accommodations for this course, please contact the Center for Access and Academic Testing, at 254.968.9400 or caat@tarleton.edu. The office is located in Math 201. More information can be found at www.tarleton.edu/caat or in the University Catalog.

Student Safety and Title IX: You are in college to achieve academic success, but you must feel safe and take care of yourself to reach your full potential. You have the right to pursue your education in a safe environment. Title IX makes it clear that violence and harassment based on sex and gender are civil rights offenses subject to accountability. If you or someone you know has been harassed or assaulted, there is help and support on campus. You may seek assistance confidentially through the Student Counseling Center or the Student Health Center. You may also make a report to the campus Title IX coordinator, which may trigger a university investigation (not a criminal investigation). Additionally, you may pursue criminal charges through the university police department. If the assault occurred away from campus, UPD can assist you in connecting with the appropriate law enforcement agency.
Student Counseling Center: 254-968-9044 (phone is answered 24 hours a day, 7 days a week), TSC 212
Student Health Services: 254-968-9271, TSC 212
Title IX Coordinator: 254-968-9754, Admin Annex 1, Room 112
University Police Department: 254-968-9002, located on the corner of Harbin and Frey

Standards of Conduct & Academic Dishonesty: Cheating, plagiarism, or doing work for another person who will receive academic credit is impermissible. This includes the use of unauthorized books, notebooks, or other sources in order to secure or give help during an examination, the unauthorized copying of examinations, assignments, reports, or term papers, or the presentation of unacknowledged material as if it were the student's own work. Disciplinary action may be taken beyond the academic discipline administered by the faculty member who teaches the course in which the cheating took place.

Tarleton State University expects its students to maintain high standards in personal and scholarly conduct. Students guilty of academic dishonesty are subject to disciplinary action. Academic dishonesty includes, but is not limited to, cheating on examination or other academic work, plagiarism, collusion, and the abuse of resource materials. The faculty member is responsible for initiating action for each case of academic dishonesty that occurs in his/her class. Academic honesty is expected. Cheating will not be tolerated and will result in automatic failure of the course. The University's Academic Integrity Policy will be maintained. Students are responsible for knowing and abiding by the policies and information contained in the Tarleton Student Handbook. [See the TSUSH]

Core Values: Tarleton State University's core values are integrity, leadership, tradition, civility, excellence, and service. Central to these values is integrity, which is maintaining a high standard of personal and scholarly conduct. Academic integrity represents the choice to uphold ethical responsibility for one's learning within the academic community, regardless of audience or situation.

Academic Integrity Statement: Tarleton State University's core values are integrity, leadership, tradition, civility, excellence, and service. Central to these values is integrity, which is maintaining a high standard of personal and scholarly conduct. Academic integrity represents the choice to uphold ethical responsibility for one's learning within the academic community, regardless of audience or situation

Academic Civility Statement: Students are expected to interact with professors and peers in a respectful manner that enhances the learning environment. Professors may require a student who deviates from this expectation to leave the face-to-face (or virtual) classroom learning environment for that particular class session (and potentially subsequent class sessions) for a specific amount of time. In addition, the professor might consider the university disciplinary process (for Academic Affairs/Student Life) for egregious or continued disruptive behavior.

Academic Excellence Statement: Tarleton holds high expectations for students to assume responsibility for their own individual learning. Students are also expected to achieve academic excellence by:

- honoring Tarleton's core values.
- upholding high standards of habit and behavior.
- maintaining excellence through class attendance and punctuality.
- preparing for active participation in all learning experiences.
- putting forth their best individual effort.
- continually improving as independent learners.
- engaging in extracurricular opportunities that encourage personal and academic growth.
- reflecting critically upon feedback and applying these lessons to meet future challenges.

Academic Affairs Service statement: Tarleton faculty, staff, and students are expected to model responsible citizenship through service activities that promote personal and academic growth while enhancing the university, local, regional, national, and global communities. These activities will foster a culture of academic/public engagement that contributes to the achievement of the university's mission and core values.

Student Responsibilities: The student is solely responsible for:

- Completing each assignment by the specified due date.
- Obtaining assignments and other materials for classes from which they are absent.

- Utilizing, as needed, all available study-aid options (including meeting with the instructor, referring to outside texts, etc.) to resolve any questions that they might have regarding homework, course material, and/or projects.
- Giving as much of an effort as it takes to pass this course.

Absence Policy: Tarleton differentiates between a failed grade in a class because a student never attended (F0 grade), stopped attending at some point in the semester (FX grade), or because the student did not pass the course (F) but attended the entire semester. These grades will be noted on the official transcript. Stopping or never attending class can result in having to return aid monies received. For more information see the Tarleton Financial Aid website.

Copyright Information: Tarleton State University is committed to adhering to all applicable laws regarding intellectual property, specifically the rights of copyright holders and compliance with copyright law. It is the responsibility of all members of the Tarleton State University community to make a good faith determination that their use of copyrighted materials is in compliance with Title 17 U.S. Code, the United States Copyright Act, Fair use, Digital Millennium Copyright Act of 1998, and the Technology, Education, and Copyright Harmonization (TEACH) Act of 2002. Guidelines in use at Tarleton State University regarding copyright can be found on the Fair Use, Copyright, and the TEACH Act Information page. For more information, please contact Ms. Jennifer Sherwood at jsherwood@tarleton.edu.

Please be aware that copyright protection also extends to the use of films for educational purposes. It is acceptable to show a full-length feature film in a face-to-face class, if the film 1) was acquired through library check out or legally purchased and 2) pertains directly to the curriculum for that class. It cannot be legally shown in its entirety in an online class or to the public. A more in-depth presentation of information can be found at: http://www.ala.org/advocacy/copyright/teachact/faq

Notes:
- In the event that the university is closed for a scheduled class time, whatever was scheduled for that day and/or whatever was due that day will be scheduled and/or due on the next scheduled class time.
- You are expected to present a TSU ID upon request.
- **All items contained in this syllabus are subject to change as the semester progresses. Students will be notified of any changes.**