

Instructor: Dr. Scott Cook	E-mail: scook@tarleton.edu
Class Time: TR 6:00-7:15PM	Classroom: Math Bldg 333
Phone: 254-968-1958	Office Hours: scheduler below
Office: Math 132	https://calendly.com/scook-10/15min
Zoom for Class Meetings https://tarleton.zoom.us/joining/97197153358	Zoom for Office Hours https://tarleton.zoom.us/my/scook

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

- Data Science Foundations with Python zyBook
 - <https://learn.zybooks.com/zybook/TARLETONMATH5364Spring2023>
 - access code posted in canvas
- Google Drive & Colab account
- GitHub account
 - https://github.com/drscook/m5364_23sp_data_science1
- Webcam & stable internet
- Canvas
 - <https://www.tarleton.edu/oiss/canvas-login.html>
- We will use a lot of open-source resources

Topics

Anytime we try to understand patterns in data, we are doing data science. This field is enormous; it is impossible to cover it in 2 semesters. The first semester introduces the data science workflow and core concepts and basic algorithms for each major step. The second semester explores advanced algorithms and rapidly surveys as many data science niches as possible. Core topics for first semester:

- Intro, terminology, big ideas, and workflow
- Probability & Statistics (review)
- Data Wrangling
- Data Exploration & Visualization
- Regression
 - Linear
 - Logistic
- Model Evaluation & Selection
 - Metrics
 - Class Imbalance & Multiclass Problems
 - Cross-validation
 - Bootstrap
 - Hyperparameter Tuning
 - Underfitting vs Overfitting
 - Curse of Dimensionality
- Unsupervised Learning
 - Dimensionality reduction
 - Principal Component Analysis

- Factor Analysis
 - Feature Agglomeration
- Clustering
 - k-Means Clustering
 - Gaussian Mixtures
 - Hierarchical Clustering
 - DBSCAN
- Supervised Learning
 - k-Nearest Neighbors
 - Naive Bayes
 - Support Vector Machines
 - Decision Trees & Random Forests
 - Ensemble Methods
- If time allows, we will introduce selected topics covered in Data Science 2:
 - Artificial Neural Networks
 - Bayesian Inference
 - Manifold Learners
 - Advance Tree-based Learners
 - Interactive Visualizations & Dashboards
 - Other topics as chosen by instructor

ZyBook

I'm excited to try the Python Data Science zyBook. It is a textbook, cloud computing platform, and homework system all in one! It embeds *participation activities (PA)*, *challenge activities (CA)*, and **zyLabs** directly into the chapters which are autograded to give you immediate feedback. I will typically assign PA to be due *before* class, CA due a few days after class, and zyLabs due about a week after class. Again, these are autograded and allow you to retry until you master it.

The zyBook is a significant course redesign inspired by prior student feedback. I'd really like your feedback on the zyBook itself and ways I can use it more effectively.

Other Assignments

In addition to the zyBook assignments, I will give some homework of my own design which I intend to grade myself. While zyBook assignments focus on code and mathematics, these assignments will ask you to write explanations, design visualizations, critically analyze results, etc. In other words, these will help you think and talk like a data scientist.

Exams

There will be one midterm and a final exam. We will wait a few weeks to schedule the midterm so we can avoid overlap with Real Analysis exams. The university calendar puts our final exam Thursday May 4 at 6:30-8:30pm, but we may reschedule it. Exams will be held in-person unless you are a remote student. Please contact me to make arrangements.

Project

You will do a capstone project using techniques from the course, due at the end of the semester. It must use both unsupervised and supervised learning techniques to perform interesting regression or classification task(s). I will post detailed instructions shortly.

Start looking for your datasets and project ideas **now**! It will take much longer than you expect to find and wrangle your data before you can start doing analytics. You will probably need to hone your questions several times before you find the “Goldilocks project” - not too simple, not too complex, just right.

Start now! The end of the semester will be stressful and demanding. We will cover necessary topics throughout the semester, so you can't finish before we cover them. But you can start. A wise student will identify datasets THIS weekend and apply techniques as we learn.

You will submit project checkpoints during the semester as follows:

- Feb 10: Data
- Feb 24 : Data + Background + Tasks
- Mar 10: Data + Background + Tasks + Data Wrangling
- Mar 31: Data + Background + Tasks + Data Wrangling + Preliminary Results
- Apr 14: Rough Draft
- Apr 28: Final Draft

You can (and should) revise previously submitted sections. Most students end up adjusting their data, tasks, and/or wrangling once they start running supervised learning and get bad results. That is typical - don't be surprised, frustrated, or concerned.

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.

PA:10%, CA:15%, ZyLabs & my assignments:25%, Project:30%, Midterm Exam:10%, Final Exam:10%
F: [0%, 60%), D: [60%, 70%), C: [70%, 80%), B: [80%, 90%), A: [90%, ∞)

Intended Outcomes

1. Appraise raw data to detect data quality issues
2. Select interesting subsets or features contained within the data
3. Assess model effectiveness
4. Transform raw data into a form appropriate for modeling
5. Select and train appropriate models using the transformed data

Communication

- Email scook@tarleton.edu - I strive to respond within 48 hours
- Office hours: You can schedule a meeting with me anytime during business hours that I am not already booked here: <https://calendly.com/scook-10/15min>. These meetings will usually be on zoom at <https://tarleton.zoom.us/j/scook>, but may occur in my office Math 132
- I post announcements on Canvas that should trigger emails to your school account. You should receive one today. If not, please [check whether you have opted out](#).

Collaboration Policy

Collaboration is extremely beneficial, especially when writing code. 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 so. You're going to get stuck and frustrated. Many of us suffer from “tunnel vision” that prevents us from seeing our own errors. Work with your classmates to help get out of the tunnel. I strongly urge you to work in the Math Building as much as possible to facilitate collaboration.

However, you **MUST** make sure you have mastered it for yourself. The exams will ask you to write code and describe concepts on your own. So, after you collaborate, rewrite the code from scratch on your own without looking at the collaborative code.

Identify collaborators on anything that I will grade (non-zyBooks assignments, project, etc). You will not lose points for having collaborators. But you might lose all credit if I can tell you collaborated and did NOT identify.

Attendance

This hybrid course is available face-to-face and via zoom, which will record and post automatically to Canvas. I strongly urge you to attend face-to-face unless you absolutely can't make it. Life presents constraints that we can't always control, but students always tell me they learn far more when they are present in the classroom. You are responsible for all content covered in class whether you are present or not.

No class: 3/14, 3/16, 3/23

Expectations, Mindset, and Time Management

Data science pulls from at least 3 different fields: math/stat, computer science, and domain knowledge. We often need to learn multiple things and multiple sources simultaneously. This means you will end up jumping around the textbook, using lots of internet resources, and collaborating with your classmates. The zyBook is great, but do not rely solely on it, my lectures, and stuff I post. You must leverage the entire internet.

Real-world projects will require you to learn new things on your own. Data science changes profoundly every few months, so you will constantly need to teach yourself new things. Get accustomed to using internet resources and good-old-fashioned trial and error.

Homework and projects usually take much longer than you expect due to code bugs. It might not be your fault - Python packages change constantly and break other packages that depend on it. Procrastination will kill you in this course because a single bug can take days to squash.

University Policies

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:

Attendance records will be kept for each class meeting. As stated in the University catalog, “the faculty member has the responsibility and authority to determine whether a student may make up assignments and/or examinations resulting from absences. Students may request makeup consideration for valid and verifiable reasons such as serious illness (such as hospitalization), death in the immediate family, or legal proceedings.”

Student Success Statement – Americans with Disabilities Act:

Tarleton State University is committed to complying with the Americans with Disabilities Act (www.ada.gov) and other applicable laws. If you are a student with a disability seeking accommodation for this course, please contact the Office of Disability Resources at 254.968.9400, disability@tarleton.edu, or visit <https://www.tarleton.edu/drt/>.

COVID-related Information:

Anyone feeling ill or exhibiting any symptoms commonly associated with COVID-19 is encouraged to stay home and monitor said symptoms and/or consult a physician. Commonly associated symptoms may be found here:

<https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>

The university reserves the right to both reduce classroom capacity and utilize remote instructional modes should either become necessary during the semester. The university will provide relevant guidance concerning both classroom capacity reductions and any changes to instructional modes directly to faculty. Faculty will communicate relevant changes to students.

Textbooks:

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.

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, harassment, and discrimination 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), Traditions North (First Floor)

Student Health Services: 254-968-9271, Traditions North (First Floor)

Title IX Coordinator: 254-968-9754, Admin Annex 1, Room 106

University Police Department: 254-968-9002, located on the corner of Harbin and Frey

Student Resources and Basic Needs:

Any student who has difficulty affording groceries or accessing sufficient food to eat each day, or who lacks a safe and stable place to live, is urged to contact the professor. In Stephenville, the Tarleton Purple Pantry is located in the lower level of the Thompson Student Center with an external door facing the Dining Hall. In Fort Worth, the Tarleton Purple Pantry is located in room 360. Request swipe entry access at www.tarleton.edu/purplepantry. Learn about additional related resources at www.tarleton.edu/TexConnect.

University Policy:

Students are responsible for knowing and abiding by the policies and information contained in the Tarleton Student Handbook. [See the Student Handbook].

Academic Conduct:

Any student guilty of academic dishonesty, cheating, or plagiarism in academic work shall be subject to disciplinary action. [See the Student Handbook] The instructor may initiate disciplinary action in and case of academic misconduct.

Standards of Conduct & Academic Honesty:

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.

Academic Honesty Statement:

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 dishonest that occurs in his/her class. Academic honesty is expected. The University's Academic Integrity Policy will be maintained.

Tarleton State University Mission, Vision, and Values:

Mission: Tarleton State University, a founding member of The Texas A&M University System, transforms generations by inspiring discovery, leadership, and inclusion through educational excellence.

Vision: Tarleton will be the premier comprehensive regional university in the nation, with a keen focus on student success, teaching, and research.

Values:

- **Excellence** – Tarleton Texans pursue greatness in all efforts, promoting a spirit of leadership, adaptability, and innovation for exceptional outcomes and a better world.
- **Integrity** – Tarleton Texans hold to the highest ethical standards and commit to serving others through transparency and accountability in everything we do.
- **Respect** – Tarleton Texans create an inclusive community by embracing diverse perspectives with civility, honoring tradition, and promoting teamwork so everyone thrives.

Notes:

- Tarleton State University will participate in Service Day on Thursday, March 25, 2022. On this day, face-to-face classes will not meet so that students and faculty can participate in a service project during that time.
- 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.
- 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.
- If the university must switch to a remote instructional mode during the semester, then access to a computer with a microphone and a webcam will be required for synchronous delivery of content and/or synchronous online testing using proctoring software.
- 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 in advance of any change.

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>