



INFO 3700/4390: Advanced Predictive Modeling with R

<p><u>Term and Credits:</u></p> <p>Summer 2023</p> <p>4 Credits</p> <p>4390-1 CRN 1769</p> <p>3700-1 CRN 1766</p>	<p><u>Location:</u></p> <p>T/Th 5:30pm – 7:40pm DCB 210</p> <p>You will be required to review some material outside of class which will be delivered through Canvas. Make sure you have a good internet connection during class for access to Canvas. Also make sure your Canvas notifications are set appropriately.</p>
<p><u>Instructor:</u></p> <p>Name: Kellie Keeling</p> <p>Department: Business Information & Analytics</p> <p>Office Hours DCB 590/Zoom: T/Th 4:30-5:25p https://udenver.zoom.us/my/kkeeling</p> <p>Make an appointment: https://bit.ly/KeelingBooking</p>	<p><u>Communication Conduct:</u></p> <p>Feel free to refer to me as Dr. Keeling, Professor, or Kellie as you feel comfortable.</p> <p>Email is usually the best way to contact me. If I haven't responded in 36 hours, feel free to resend your message.</p> <p>I will send class level communications via Canvas announcements. I will typically initiate communication with individual students directly through your DU email or through Canvas email.</p> <p>Email: kkeeling@du.edu Office Phone: 303-871-2296 (forwards to my cell)</p>

COURSE DESCRIPTION:

This course serves as an introduction to advanced predictive modeling and statistical learning using the R statistical software. Specific topics include linear, non-linear, and logistic regression, classification, resampling methods, and non-linear regression, tree-based methods, support vector machines, and neural networks. The students will learn how to communicate their results (business reports, dashboards, etc.) of the various modeling exercises and projects using R Studio and Quarto Markdown files.

PREREQUISITES/CO-REQUISITES:

Prerequisites: INFO 4300 or INFO 3200

LEARNING OUTCOMES:

Principal Content Elements:

1. R programming knowledge including cleaning, manipulating, and analyzing data.
2. Advanced business predictive modeling and statistical learning techniques and tools using R and RStudio
3. Business presentation techniques including reports and dashboards using R and RStudio.



PROGRAM LEARNING OUTCOMES:

1. **Ethical Decision-Making** - Graduates will recognize and consider the legal and ethical implications of business decisions.
2. **Technical Knowledge** - Graduates will possess basic knowledge of the functional areas of business and how they integrate to achieve organizational goals.
3. **Communication and Interpersonal Effectiveness** - Graduates will communicate effectively and have the interpersonal skills necessary to collaborate effectively in a business environment.
4. **Critical Thinking** - Graduates will apply critical thinking skills to identify and resolve business issues.
5. **Global/Intercultural Competence** - Graduates will be prepared to meet the challenges and opportunities of intercultural differences that impact business organizations in a global society

STUDENT LEARNING OUTCOMES:

1. Apply predictive modeling and statistical learning methodology to a variety of business problems using R and RStudio.
2. Apply innovative, interactive data visualization tools/methods to both the raw data and model results as a means for communicating business insights.
3. Derive a set of conclusions and business recommendations from these models.
4. Learn how to effectively deliver the results of the modeling exercises and projects in technical and non-technical manners.
5. Continue to develop the critical thinking and problem solving skills necessary to produce a business decision or recommendation from a raw data set.

REQUIRED MATERIALS:

1. An Introduction to Statistical Learning with Applications in R, Publisher: Springer. First or Second Edition. You can get a digital copy from the bookstore and/or it is also available online for free at <https://www.statlearning.com/>
2. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, ISBN 978-1491910399, Publisher: O'Reilly. (Note: This book is available online for free at <http://r4ds.had.co.nz/> .)
3. The R statistical software is often referred to as the lingua franca of data science. As such, we will rely extensively on this software, along with RStudio integrated development environment (IDE). Please download both R and RStudio prior to the first day of class or make sure you have gotten into the R Studio Server. Both R and RStudio are available for Windows, Macs, and Linux computers.

POLICIES AND RESOURCES:

Class Policies. (How to be successful in this course)

Classroom Attendance: Please plan to join us in person in for class unless you have an excused absence (illness, travel for work/class/capstone). I will excuse 1 day during the quarter for personal travel. I will plan to record the class on Zoom if you miss for an excused absence.

Getting Help: If you are having difficulty with the course material, please contact me at your earliest convenience – particularly if you feel you are getting behind - That's OK! It happens! Don't delay because you feel embarrassed. I get behind and overwhelmed myself! My calendar is tied to my **Bookings appointment page** (<https://bit.ly/KeelingBooking>) and I have availability on most Saturday mornings and Sunday afternoons. If you are having an emergency situation, please let me know so we can plan accordingly. If you contact me with questions, I plan to respond within 36 hours.

Individual Work: For individual work, you may discuss homework so long as **each person is using their own technology to solve problems and turning in their individual work. Looking over someone's shoulder or watching as they do work, no matter how much you feel like you are participating, is not working together.** Turning in individual work done by

someone else is academic dishonesty and can result in penalties ranging from **zero** points on the assignment to an "F" in the class and dismissal from the class.

Late Work: You will be given the opportunity for **1 Homework Oops Ticket, 1 Project Part 1 or 2 Oops Ticket, and 2 Group Task Oops Tickets** which will allow you turn in late work without an excuse. **The extension with an Oops ticket is 1.5 weeks unless the quarter ends first.** To request an Oops ticket – put a comment with your submission on Canvas of the form **"I request 1 of 1 HW Oops Ticket" or "I request 1 of 2 Group Task Oops Ticket"**.

1. Everything should be turned in by the indicated due dates on Canvas. But if it is due by 11:59pm, I won't count it late as long as it is in by 8a the following morning. You might want to start early just in case you have a computer failure.
2. **Late work will be accepted with a penalty of 12 percentage points a day (.5% per hour)** unless you have requested an Oops ticket. The Final Part of the Project must be submitted/presented on time.

GRADING STRUCTURE, SCALE, AND POLICIES:

A: 93-100%; A-: 90-92.9%; B+: 87-89.9%, B: 83-86.9%; B-: 80-82.9%; etc.

Student performance will be evaluated and calculated on the items below:

ASSESSMENTS:

Activity	Percentage
Pre Class: DataCamp 7 Chapters	7%
Pre Class: Reading/Lecture Quizzes	20%
In Class: Group Task Mini Assignments & Check In Quizzes (6)	20%
Homework (3)	23%
Project (3 parts)	30%
Total	100%

DataCamp Chapters: At the beginning of the course, there will be 7 DataCamp chapters assigned to help everyone come up to speed with the R statistical programming language. Sign up for the class "Adv Predictive Analytics w R Su23" (<https://app.datacamp.com/groups/adv-predictive-analytics-w-r-su23/>) through the email you should have received or with the link below. I will automatically be able to see your progress on these assignments. If you completed them previously, you do not have to repeat them. But you do need to upload screenshots to the assignment.

Reading/Lecture Quizzes: Students are expected to read and study the assigned textbook pages in addition to watching lecture videos before class. Topics that are not covered in class, but are still in the text are eligible to appear on assignments. Reading is encouraged prior to class to improve learning. To help motivate you, there will be quizzes posted in Canvas that will be due before class.

Kurzweil 3000 - <https://otl.du.edu/knowledgebase/introduction-to-kurzweil-3000/> Note if you would like to have software "read" the textbook (or other web pages or PDFs) to you, you have access as a student to Kurzweil 3000.

Homework: There will be 3 homeworks posted throughout the quarter.

Project: The focus will be on data analysis and a dashboard presentation to your peers using the skills obtained in this class. The project will be broken into 3 phases (initial proposal, initial data exploration and dashboard creation, and final dashboard and presentation).



Educational Assessment

The Daniels College of Business may use assessment tools in this course and other courses for evaluation. Educational Assessment is defined as the systematic collection, interpretation, and use of information about student characteristics, educational environments, learning outcomes and client satisfaction to improve program effectiveness, student performance and professional success.

Group Tasks: Each class there will be at least one short assignment to complete to review the material from the reading and in class notes. You will submit your assignment at least 2 hours before the following class. I encourage you to work with others during our class time on these assignments. EVERYONE in the group submit. These will count as participation credit based on attempting the work. If you did not attempt all the questions with a good faith effort, I will score down accordingly. If you are absent for class, you can complete them on your own.

100%	90%	80%	0%
In Class: On Task Entire Time Out of Class: Completed All	Completion Acceptable	Completion OK	Other

NOTE: Once you submit you can see my solution to the group task under Modules. Use this to **self-grade** your work and **FIX ANYTHING THAT DID NOT WORK BEFORE** our next class. We will go over some of the answers at the beginning of the following class.

In Class Check-ins: During some classes we will have short individual quizzes (6 total) to check your understanding of the material at that point in the quarter. You must be present to take them. They can be made up within 1 week for an excused absence. They will be announced before class.

Syllabus Policy: This syllabus is subject to change based on the needs of the class, and at the discretion of the instructor.

UNIVERSITY EXPECTATIONS, POLICIES, AND RESOURCES:

Students with Disabilities. A student who qualifies for academic accommodations because of a disability must submit a Faculty Letter to the instructor from the DU Disability Services Program (DSP) in a timely manner, so that the needs of the student can be addressed. Accommodations will not be provided retroactively, e.g., following an exam or after the due date of a project. DSP determines eligibility for accommodations based on documented disabilities. DSP is located in Driscoll Center South, Garden Level, Suite 22 (303-871-3241).

Inclusive Learning Environments.

- In this class, we will work together to develop a learning community that is inclusive and respectful. Our diversity may be reflected by differences in race, culture, age, religion, sexual orientation, socioeconomic background, and myriad other social identities and life experiences.
- The goal of inclusiveness, in a diverse community, encourages and appreciates expressions of different ideas, opinions, and beliefs, so that conversations and interactions that could potentially be divisive turn instead into opportunities for intellectual and personal enrichment.
- A dedication to inclusiveness requires respecting what others say, their right to say it, and the thoughtful consideration of others’ communication.
- Both speaking up and listening are valuable tools for furthering thoughtful, enlightening dialogue. Respecting one another’s individual differences is critical in transforming a collection of diverse individuals into an inclusive, collaborative and excellent learning community.
- Our core commitment shapes our core expectation for behavior inside and outside of the classroom. Office of Diversity, Equity, and Inclusion website (<https://www.du.edu/diversity-inclusion/index.html>).



University Expectations. Please review the University Expectations on the Daniels College of Business syllabus webpage (<http://daniels.du.edu/university-expectations/>)

- University of Denver Honor Code
- Policy Concerning Official Communication
- Students with Disabilities
- Policy Concerning Religious Accommodations
- Policy Concerning Emergency Procedures
- Policy Concerning Conflicts of Interest, Including Gifts from Students

Tentative Schedule:

		TUE	THU
Week 0	Complete DataCamp Courses (Intro to Tidyverse/Intermediate R)	BEFORE CLASS STARTS	
Week 1	R/R Studio Intro/Tidyverse and Control Logic/ Statistical Learning	Jun 20 (Day 1)	Jun 22 (Day 2)
Week 2	Simple/Multiple Linear Regression / Regression Assumptions & Transformations	Jun 27 (Day 3)	Jun 29 (Day 4 & 5) CI1
Week 3	Logistic Regression	Jul 4 (Holiday)	Jul 6 (Day 6) CI2
Week 4	Discriminant Analysis / Cross Validation / Bootstrap Resampling	Jul 11 (Day 7)	Jul 13 (Day 8)
Week 5	Ridge Regression and Lasso / Smoothers: Polynomials and Splines	Jul 18 (Day 9) CI3	Jul 20 (Day 10)
Week 6	Gen Additive Models / CART Trees	Jul 25 (Day 11)	Jul 27 (Day 12) CI4
Week 7	Bagging/Random Forest/Boosting Models	Aug 1 (Day 13)	Aug 3 (Day 14) CI5
Week 8	Support Vector Machines / Neural Networks	Aug 8 (Day 15)	Aug 10 (Day 16)
Week 9	Deep Learning / Present Projects	Aug 15 (Day 17) CI6	Aug 17 (Day 18)