

# STAT 5533 - Data Mining & Statistical Learning

TR, 4:30 - 5:45 pm, MCS 109

Spring 2025

Instructor: Dr. Tyler Cook

Email: [tcook14@uco.edu](mailto:tcook14@uco.edu)

Phone: (405) 974-5620

Office: STEM 240

Office hours: TR 11:00 AM - 12:00 PM and 2:00 - 4:00 PM,

Other times by appointment

**Course Description:** This course will expose students to modern data analysis techniques for regression, classification, and unsupervised learning. Students will apply machine learning algorithms to real-world data using the R programming language. Selected topics include: penalized regression, regression splines, decision trees, support vector machines, principal components analysis, and clustering.

**Prerequisite(s):** STAT 4213 with a minimum grade C or permission of the instructor.

## Course Material:

- **Text:** *An Introduction to Statistical Learning with Applications in R*, by James, Witten, Hastie, and Tibshirani (2nd edition). A free PDF copy of the textbook is available on the authors' website: <http://www.statlearning.com>.
- **Software:** All programming and data analysis will be done using R. R can be downloaded from <https://cran.r-project.org/>. In class demonstrations will be done using RStudio which is available at <https://www.rstudio.com/>. Code used in class can be downloaded from <https://github.com/tylerlewiscook/UCO-Data-Mining-Stat-Learning>.

## Course Objectives:

Upon completion of the course, the student will be able to:

1. Differentiate and choose between a number of statistical methods for the analysis of supervised and unsupervised learning problems.
2. Perform common statistical analyses in R.
3. Develop plans for analyzing real-world data.
4. Critically interpret, summarize, and communicate the results of a data analysis problem.

## Course Requirements and Evaluation:

- **Homework:** A total of eight homework assignments will be assigned throughout the semester. Assignments will mostly consist of data analysis using R, and code will be submitted online for grading. A template for homework submissions can be found on D2L. All code and interpretation of results must be your own. Late homework will not be accepted.

- **Exams:** A total of three exams will be administered in class during the semester. Students will be allowed to use a calculator, provided formulas and statistical tables (as needed), and a laptop running R for all exams. All exams must be taken at the scheduled time. Exam dates: February 18, March 25, and May 1.
- **Make-up Policy:** No make-up exams will be given. If a student must miss an exam for an emergency, then their final exam will be weighted to account for the missing midterm. Some form of documentation must be provided to the instructor as soon as possible in order to verify the absence.
- **Project:** Each student will complete a group project where a real-world problem will be analyzed from start to finish (data acquisition, cleaning, analysis, and reporting). Various parts of the project will have due dates throughout the semester, and the final product will be a written summary and a presentation at the end of the semester. Presentations will be given during finals week at the time scheduled by the university for the final exam (May 8 at 5:30 PM).
- **Point breakdown:** The course will have a total of 500 available points. Each homework assignment will be worth 25 points with the lowest homework score dropped, the midterm exams will each be worth 75 points, the final exam will be worth 100 points, and the project will be worth 75 points.
- **Graduate Students:** Students enrolled in STAT 5533 will be required to complete extra questions on homework assignments and exams.

#### Grade Distribution:

A	B	C	D	F
90-100%	80-89%	70-79%	60-69%	<60%

**Students with disabilities:** The University of Central Oklahoma complies with Section 504 of the Rehabilitation Act of 1973 and the Americans with Disabilities Act of 1990. Students with disabilities who need special accommodations must make their requests by contacting Disability Support Services at (405) 974-2516. The DSS Office is located in the Nigh University Center, Room 305. Students should also notify the instructor of special accommodation needs as soon as possible as accommodations are proactive, not retroactive.

**Academic Honesty:** All students are expected to follow the UCO Student Code of Conduct. While discussing assignments with classmates is acceptable, it is important to ensure that the work turned in is your own. At minimum, any student found guilty of cheating will be given a zero on that particular assignment/exam.

#### Miscellaneous:

- The Student Information Sheet and Syllabus attachment can be found by going to <http://www.uco.edu/academic-affairs/files/aa-forms/StudentInfoSheet.pdf>.