

# DSCI445: Statistical Machine Learning

Fall 2025

<http://dsci445-csu.github.io>

Dr. Andee Kaplan

Lectures/Labs: TTh 8:00am - 9:15am Natural Resources Building 109

Office Hours: Th 12:30pm - 2:30pm Statistics Building 208

[andee.kaplan@colostate.edu](mailto:andee.kaplan@colostate.edu)

## Course Objectives

Statistical Learning refers to a set of tools for modeling and understanding complex datasets. The area combines knowledge from statistics and computer science to tackle these “big data” problems and has become a popular area of work.

By end of course, students will be able to:

1. Formulate prediction problems as statistical machine learning problems (classification, regression, clustering).
2. Choose and apply the appropriate learning methods to their data.
3. Conduct well thought-out statistical machine learning experiments and interpret the results.
4. Write technical reports describing their work.

## Prerequisites

DSCI 320, DSCI 369 and STAT 341.

## Texts

An Introduction to Statistical Learning with Applications in R (2017) by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani – Available free [here \(https://static1.squarespace.com/static/5ff2adbe3fe4fe33db902812/t/6009dd9fa7bc363aa822d2c7/1611259312432/ISLR+Seventh+Printing.pdf\)](https://static1.squarespace.com/static/5ff2adbe3fe4fe33db902812/t/6009dd9fa7bc363aa822d2c7/1611259312432/ISLR+Seventh+Printing.pdf).

Optional Reference: The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2009) by Trevor Hastie , Robert Tibshirani, and Jerome Friedman – Available free here: <https://web.stanford.edu/~hastie/ElemStatLearn/>

## Computing

We will use RStudio (<https://rstudio.com>), R (<https://r-project.org>), GitHub (<https://github.com>). All software is free and open source.

**Please install** on your own computer or use the class RStudio Server (details to follow).

For your homeworks, you are free to use any other language you like, but I may not be able to help you with your computing if the language is unfamiliar to me.

## Classwork and Grading

All graded classwork must be fully **reproducible** and **readable** by the instructor and TA. In other words, we need to be able to run your code and have it produce the product you turned in and your turned in document must be legible by the grader. If this is not the case, it will be reflected in the grading. A copy of your homework will need to be turned in to <https://canvas.colostate.edu> and the corresponding document used to generate your homework will need to exist on the server for full credit.

**Labs (5%)** Labs will be completed periodically in class. You will turn in what you completed in class on Canvas, due at **9:20am on lab days**. These will be graded for completion, not correctness.

**Homework (65%)** Homework will be assigned bi-weekly. All homework assignments are due at **11:59pm on the due date**. Each homework assignment will receive equal weight in the final grade and the one lowest homework assignment grades will be dropped. Late work is not accepted except in rare cases (see Documented emergencies below).

**Project (30%)** There will be a final project that will consist of an analysis of real data using the tools learned in class (this can include participation in an online data science competition, e.g. Kaggle).

You will write a paper and give an in class presentation. More details will be announced later.

Grades will be assigned according to the following intervals:

A	A-	B+	B	B-	C+	C	D	F
[100, 93]	[93, 90]	[90, 87]	[87, 83]	[83, 80]	[80, 77]	[77, 70]	[70, 60]	[60, 0]

Any grading dispute must be submitted in writing to me within one week after the work is returned.

**Extra credit** Any extra credit will be announced in lecture only. If you miss lecture, you *may* miss chances for extra credit.

## Policy Regarding Academic Honesty

Statisticians and Data Scientists need to have high ethical standards. Thus, I expect each of you to hold high ethical standards and to act with academic integrity in this class. If you have questions about what integrity means, please feel free to ask me. Behavior that will not be tolerated in this class includes turning in a copy of somebody else's homework or code as your own, copying from somebody's exam, or failure to cite sources.

This course adheres to the CSU Academic Integrity Policy as found on the Students' Responsibilities pages of the CSU General Catalog in the Student Conduct Code. Violations will result in zero points for the assignment as a minimum penalty. In addition, CSU policy requires instructors to report violations to CSU's Office of Conflict Resolution.

## Policy Regarding use of LLMs

All assignments should be fully prepared by the student. Developing strong competencies in the skills associated with this course will prepare you for success in your degree pathway and, ultimately, a competitive career. Therefore, the use of LLMs (i.e., generative AI tools) to complete any aspect of assignments for this course are not permitted and will be treated as plagiarism. If you have questions about what constitutes a violation of this statement, please contact me.

## Documented Emergencies

If you have a problem that will require you to miss a due date, please discuss this with me in advance if possible. I can grant a rare exception when the reason relates to severe and unavoidable medical or personal emergency. Documentation will be required. Things that typically are not an emergency: vacation, family reunions, ordinary work commitments, job seeking, or other voluntary events. Please schedule these so that they do not conflict with your classes.

## **Support Services Available**

CSU provides policies relevant to your courses and resources to help with various challenges you may encounter. Please see <https://col.st/2FA2g> for more details.

## **Disclaimer**

I reserve the right to make amendments to the syllabus and the schedule throughout the semester. Any updates will be posted on the class website and announced via e-mail and in class.