CS 300: Data Mining and Machine Learning Fall 2025

This course considers methods concerned with the analysis of date to draw conclusions. Rather than simply using existing libraries or tools ("black boxes"), we'll examine them more closely to understand how they work, and how modifying them influences their behavior. In this way, you will ideally be better able to understand the details of several machine learning algorithms, and be better qualified to understand their advantages and drawbacks.

Meetings: TR 2:40–4pm in SCSC 389.

About half of our class meetings will take place in this classroom, and the other half will occur in SCSC 377, where we'll be working labs on the computers. If you own a laptop computer, please bring it with you to **all** class meetings.

Instructor: R. Scott Linder, SCSC 386, rslinder@owu.edu

Office Hours: M W 1–2 and 4–5; R 4–5; F 4–5.

You can also drop in any other time I'm in my office with the door open.

Prerequisite: CS 110 and CS 210. You should be comfortable working with numbers and probability, and you should be familiar with basic calculus knowledge. C++ programming skills you've developed will make programming in R more accessible.

Textbook: *Data Mining: Practical Machine Learning Tools and Techniques*, 4th edition, by Witten, Frank, Hall and Pal.

Blackboard: Electronic copies of assignments, labs and programs shown in class will be posted on the Blackboard site. I will also post announcements there. I will assume that you have seen all announcements, so it's your responsibility to check the Blackboard site regularly (at least two or three times per week).

Course Grade: Your total course average (TCA) will be computed using the following weights.

Assignment average	25%
Exam 1 (T 10.7)	20%
Exam 2 (T 11.18)	20%
Group Project	20%
Attendance:	15%

Your course grade will depend only on your TCA. The grade scale mapping your TCA to a letter grade will not be determined until the semester ends. However, if your TCA is at least 90%, your grade will be *at least* A–, if it's at least 80%, your grade will be *at least* B–, and it it's at least 70%, your grade will be *at least* C–. Any student with a TCA lower than 50% will fail the course (F), while any student with a TCA between 50% and 60% cannot earn a course grade higher than D.

Assignments: You should expect to work assignments, typically submitted every 1.5 weeks (2 assignments per 3 weeks). All of them will require R programming. Unless told directly by the instructor, you must work assignments **independently***, with no help from fellow students or from an AI resource. Assignments are due by midnight of the due date. Late assignments will be assessed a 10% penalty every 12 hours. For example, if you turn the assignment in 6 hours late, there won't be a penalty, but if you turn it in at noon (12 hours late), you'll lose 10%, and so on. Your lowest course assignment score will be dropped.

You may be asked to meet with the instructor to discuss technical aspects of any assignment you turn in. Here you will need to demonstrate that you understand completely the work you submitted. References to AI or YouTube will result in a severe assignment grade penalty.

*To make this clear: If you receive (or provide) material in the form of any code, diagrams, pictures, algorithms or formulas from (or to) anybody or from any AI resource, then you are not working independently. Asking AI to generate an algorithm or code for you is a form of academic dishonesty. So is exchanging assignment work of any kind with fellow students.

Group Project: Students will work in groups of 3 on a substantial project that integrates course concepts. Groups will present their work to the class at the end of the semester. Your grade on this group project will depend heavily on meeting milestone checkpoints throughout the semester. More detail about this project will be provided later.

Attendance: Each missed class will result in a 10% penalty to your credit for this component of your course grade. For example, if you miss two classes, you'd lose 20% for this component, which would reduce your TCA by 1% (20% of 5%).

General Advice:

This is an upper-level Computer Science course with a substantial math/statistics component. You're expected to spend at least 8–10 hours each and every week focused* on this class outside of class time. Maybe you'll need more than this much time, however… and it's your responsibility to adjust accordingly. If you're honestly spending 15 or more hours per week in true focus*, but are still struggling, please talk with me.

*So, what is focus? Focus means that you're spending *many* hours in real concentration (no phone or video or music distraction), working on assignments, reading the textbook, studying your notes, etc. It means that you're trying and failing many times to get your code working right before finally succeeding. It requires perseverance and persistence, and a little grit. It means that you aren't staring problems and quickly running for help the first time you hit an obstacle. Expect frustration and learn to overcome it.

Schedule (tentative):

R 8.21:	Class introduction; R introduction (Installation and Data types)
T 8.26 and R 8.28:	R introduction (Importing files, data frames, Tidyverse); Lab 1; Last day to drop the class.
T 9.2 and R 9.4:	R details (Strings, expressions, functions); Lab 2
T 9.9 and R 9.11:	R details (Merging and control structures); Lab 3
T 9.16 and R 9.18:	R details (apply, lapply, sapply functions, control structures); OWU Connection Conference
T 9.23 and R 9.25:	1-rule (1R) classification; Lab 4
T 9.30 and R 10.2:	Naïve Bayes classification (multinomial); Lab 5
T 10.7 and R 10.9:	**Exam 1**; Fall break
T 10.14 and R 10.16:	Naïve Bayes classification (gaussian); Lab 6
T 10.21 and R 10.23:	Decision trees; Lab 7
T 10.28 and R 10.30:	K-nearest neighbors; Lab 8: K-nearest neighbors
T 11.4 and R 11.6:	Linear regression; Lab 9
T 11.11 and R 11.13:	Logistic regression; Lab 10
T 11.18 and R 11.20:	**Exam 2**;
T 11.25 and R 11.27:	No Class (Thanksgiving Break)
T 12.2 and R 12.4:	Group project presentations