Machine Learning for Social and Behavioral Research

PSY 40122 | Fall 2025

University of Notre Dame

Course Information

Pre or Co-Requisites: PSY 30100 or equivalent

Target Audience: Psychology majors (seniors/juniors)

Meeting Pattern: MW 11am-12:15pm

Instructional Format: in-person

Classroom or Location: Corbett Family Hall E404

Instructor Information

Name: Oscar Gonzalez, PhD (he/him/his)

Email Address: ogonzal1@nd.edu

Office Location: Corbett Family Hall E462

Office Hours: W 10-11am, Th 2:30-3:30pm

Teaching Philosophy: This class is inherently a statistics course. Most students have had bad experiences with statistics, so they expect an uphill battle. My teaching philosophy addresses these issues by breaking concepts down as best as I can and focus on the overall messages of the course rather than in tedious equations. As an underrepresented faculty member, I try to make sure that everyone feels included and I welcome suggestions to make the classroom environment work for everyone. My background is similar to students in the **Building Bridges** program and **Balfour-Hesburgh** scholars.

I believe in you! I want to see you succeed!

Course Content

Course Description

In this day and age, we interact with many data collection tools. From loyalty cards, movie recommendations, or taking driving directions from a GPS, we are leaving a data footprint almost every day. Machine learning algorithms could help us go from raw datasets to valuable information. Machine learning has recently emerged as a major area of statistical research and is making its way into psychology. This course is an introductory seminar on the theory and application of machine learning to data analysis.

Course Texts & Materials (FREE)

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An introduction to statistical learning with applications in R* (2nd Edition). New York, NY: Springer. (ISL in tentative schedule)

Download: https://www.statlearning.com/

Supplemental readings will be posted regularly on Canvas. Citations for a few of the readings is at the end of the syllabus.

Statistical Software (FREE)

In this class we will use the R statistical language for data analysis. R is a popular, script-based statistical language that you can download for free in your own laptop. We will also use RStudio as a GUI for R, although not necessary.

Download: https://www.r-project.org/

Also, we will use generative AI, such as ChatGPT and Google Gemini, in some activities and assignments.

You can access the ChatGPT at chatgpt.com and Google Gemini Pro (with NetID) at gemini.google.com

General AI policy

Unless given permission to use those tools, each student is expected to complete each assignment without substantive assistance from others, including tools such as large language models (ChatGPT and Gemini) and AI editing tools (Grammarly). Representing work that you did not produce as your own, including work generated or materially modified by AI, constitutes academic dishonesty. **Use of generative AI in a way that violates an instructor's articulated policy or using it to complete coursework in a way not expressly permitted by the faculty member, will be considered a violation of the Honor Code.**

Class Expectations

Attendance and Accessibility

Attendance is mandatory (except when there are documented university-excused absences). I work with students to accommodate different life circumstances. Please communicate with me over email or in person about your learning needs or anything that you think can hinder your learning in class. I am here to help you.

General day-to-day structure

Most days I will lecture on a particular machine learning algorithm, discuss its theoretical underpinnings, and demonstrate some code on how to carry it out. After we finish covering an algorithm, there will be dedicated days where we have a lab session where I illustrate how to carry out the algorithm in the R statistical environment. Students will then conduct the analysis by themselves, write a short report, and turn it in for a grade. Towards the end of the semester, I will lecture on special topics in machine learning while students prepare for their final projects.

In-class expectations

Generally, I expect you to be engaged with the lecture, participate in activities, and provide minimal distractions. Some form of personal device, whether laptop, tablet, or phone, will be needed to get the most out of our class sessions. Throughout, we will be using the internet to enhance our discussions and collaborate with one another. I ask that you refrain from activities that are unrelated to the class, such as checking your email or browsing, as they can be distracting not only to yourself but also to others and me. However, I understand that emergencies happen. If you need to take care of personal matters (I hope everything is alright!), please step into the hallway.

Out-of-class expectations

Generally, I expect you to complete the reference readings listed in the syllabus prior to coming to class. If I also assign videos or other pre-class material, I expect you to complete it. This helps facilitate participation and learning the material. You can expect from me prompt replies to emails.

Learning Goals

Upon successful completion of the course, you will be able to:

1. Assess and critique the implementation of machine learning algorithms in applications in society.
2. Compare and contrast the explanatory and predictive approaches to data analysis.
3. Implement the R statistical environment to predict continuous and categorical outcomes using machine learning.
4. Communicate orally and in writing to broad audiences the steps to conduct a machine learning algorithm and when to use them.

You will meet the learning goals by engaging with the lectures and a variety of assignments, ranging from data analysis, reflection papers, ethics presentation, and final class project.

Course Assignments & Assessments

Final Project (30% of total grade)

In collaboration with other students, you will take on a comprehensive analysis of a dataset using the algorithms discussed in the class and carried out during the lab sessions. The project will consist of submitting an analytic plan, data analysis, and a final presentation of the results. The final presentation will constitute the "final exam" for the class and will take place the last day of class.

Lab Write-ups (40% of total grade)

There will be five lab write-ups where you report your findings from the lab session. Length of the report will vary but will be at least 2 pages. Lab reports are due, roughly, 3-7 days after they are assigned.

Assignment #1 (8% of total grade)

This assignment will be on a basic introduction to R.

Assignment #2 (8% of total grade)

This assignment will be on estimating the prediction performance of tree-based methods.

Assignment #3 (8% of total grade)

This assignment will be on estimating the prediction performance of k-nearest neighbors and naïve Bayes.

Assignment #4 (8% of total grade)

This assignment will be on estimating the prediction performance of linear regression methods.

Assignment #5 (8% of total grade)

This assignment will be on estimating the prediction performance of variable selection methods and logistic regression.

Ethics Presentation (10% of total grade)

There will be a group presentation on algorithm bias. The groups will find a case study in which machine learning has given unfair decisions. Students will also watch the other presentations and reflect on what they learned.

Participation (20% of total grade)

There will be several in-class activities, out-of-class activities, and discussions that will be used to assess participation.

Activities (10% of total grade)

There will be several reflection papers and written activities assigned throughout the semester. These are short tasks that are "turned in," and assigned as needed.

Discussion/Engagement (10% of total grade)

Students will be asked to participate in discussions during class and provide opinions/impressions of several topics in the class. Regular attendance helps this component, and excessive absences hurt this component. These are typically tasks that are NOT "turned in."

*****Please submit all documents in 12pt Times New Roman, double-spaced. Please use in-text citations and references in the current APA style. Match the style as close as you can.***

GRADING

Important Dates (Tentative)

	Assigned	**Due**
Reflection paper #1	08/25	08/27
Lab report #1	09/01	09/08
Reflection paper #2	09/10	09/15
Lab report #2	09/17	09/24
Lab report #3	09/29	10/06
Lab report #4	10/13	10/17
Lab report #5	11/05	11/12
Analytic Plan (in class)	11/17	11/17
Ethics Presentation	11/17	11/24
Final Presentation	12/01	12/10

Late Work

Extensions are provided for students with university-excused absences.

Life clause: Life Happens. We all lead densely-layered lives; therefore, one of my core values is leading with grace. As a result, I institute a "Life Clause:" should you need it, you may invoke the "Life Clause" on ONE of the five lab reports or reflection papers and get 2-day extension: no explanation required.

If your case is complex, know that I am committed to work with students to accommodate different circumstances that prevent them from turning in their work and provide extensions when I see fit. When there is lack of communication, here are some general guidelines on what you can expect:

Assignments: Generally, 10% off each day that it is late. Not accepted after 3 days.
Participation: Generally, these points cannot be made up.

Grading Rubrics

Assignments are graded out of 100. Each question is worth a certain number of points. The final project is graded out of 100 based on several principles, such as completion, accuracy, presentation skills, clarity, and interpretation.

Total participation points are determined by two components: half from the activities and half from discussion. The discussion points are based on two impression ratings by the instructor at each half of the semester, ranging from 0-4. Definitions are the following: 4 = regular attendance and active participation, 3 = regular attendance but some participation, 2 = regular attendance but no participation, 1 = intermittent attendance, 0 = null attendance.

Grading Scale

Adding the percentages from each component above will yield the final grade. I round up to the nearest whole number (e.g., a 94.0001 = 95).

Numeric Grade (%)	Letter Grade
95 and above	A
91 – 94	A-
88 – 90	B+
85 – 87	B
81 – 84	B-
78 – 80	C+
75 – 77	C
71 – 74	C-
61 – 70	D
60 and below	F

Grade Appeal Process

If you have any concerns with grading and/or feel you have been awarded an incorrect grade, please discuss it with me as soon as possible. Regrading is possible, but this is rare. Resubmissions or revisions of work is not allowed. If we cannot resolve the issue, you may talk to our director of undergraduate studies or department chair.

Honor Code Statement

Academic integrity is required, and I will not tolerate academic dishonesty in any form. The University's Honor Code (http://honorcode.nd.edu/) reminds our community of our shared purpose both within the institute of academia and as members of a broader humanity; the statement also outlines policy violation procedures. Any questions regarding academic integrity, particularly regarding assignments in this course, should be directed to the instructor. You will need to adhere to the following guidelines in our class: AI-generated submissions are not permitted (unless explicitly instructed to do so) and will be treated as a violation of academic integrity. To better understand Notre Dame's position on the use of large language models and related technology please consult the University's most recent statement (May 2023) regarding generative AI.

Course Schedule

The instructor reserves the right to make changes to the syllabus including project/assignment due dates.

Unit 1: Introduction to Machine Learning

Class (Date)	Topic(s)	Readings
08/25	Why Machine Learning?	Witten et al., (2016) Ch.1
08/27	On Learning and Classifying	ISL p.129-133
09/01	**Lab: Introduction to R**	**ISL p.42-51**

Unit 2: Classification Methods

Class (Date)	Topic(s)	Readings
09/03	Classification Trees with CART	ISL p.327-335
09/08	**CLASS CANCELLED**	
09/10	CART & Machine Learning Concepts	ISL p.335-340
09/11-09/15	VIDEO : Random Forest & Ensemble Methods	ISL p.340-345
09/15	Evaluating Classification Accuracy	Zou, et al. (2007)
09/17	**Lab: CART and Ensemble Methods**	**ISL p.353-359**
09/22	K-Nearest Neighbors (KNN)	ISL p.39-42; Wu et al. (2008)
09/24	Naïve Bayes	ISL p.154-158; Wu et al., (2008)
09/29	**Lab: KNN and Naïve Bayes**	**KJ (2013), Ch.13; ISL p.203-209**

Unit 3: Regression Methods

Class (Date)	Topic(s)	Readings
10/01	Multiple Regression	ISL p.59-82
10/06	Regression with Discrete Predictors	ISL p.83-87
10/08	Assumptions of Regression	ISL p.92-103
10/13	**Lab: Regression**	**ISL p.110-119**
10/15	Regression with Interactions	ISL p.87-92
10/20	**FALL BREAK**	
10/22	**FALL BREAK**	
10/27	Logistic Regression for Categorical Outcomes	ISL p.133-139
10/29	Regularization	ISL p.237-251
11/03	Forward Selection in Regression	ISL p.225-237
11/05	**Lab: Model Selection and Logistic Reg**	ISL p.171-177; p.267-279
11/10	Regression Trees for Continuous Outcomes	ISL p.328-335

Unit 4: Special Topics + Final Prep

Class (Date)	Topic(s)	Readings
11/12	Explanation and Prediction	Yarkoni & Westfall (2017)
11/17	Unsupervised Learning	ISL p.497-530
11/19	**Ethics Presentation Workday**	
11/24	**Ethics Presentation View**	
11/26	**THANKSGIVING BREAK**	
12/01	Intro to the Final	
12/03	**FINAL Lab: Stacking Algorithms**	**Witten et al. (2016), Ch.12**
12/08	Wrap-up	
12/10	**FINAL PRESENTATIONS**	

Table a: Unit/Module 4 Schedule

Reference Readings

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning with applications in R.* New York, NY: Springer

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26). New York: Springer. **KJ**

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... & Zhou, Z. H. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, *14*, 1-37.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*, 1100-1122.

Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, *115*, 654-657.

General Policies

Major Emergencies

In the event of an emergency situation or disruption to normal campus operations (weather, health emergency, etc.), our class will likely shift to remote instruction using Zoom, as well as other synchronous and asynchronous course materials. Any changes will be communicated through email and/or Canvas.

Title IX

The University of Notre Dame provides services for those who have been affected by sexual assault, sexual misconduct, dating or domestic violence, stalking and any conduct that creates a hostile environment. **For help and further information including contact information for on and off-campus resources**, please consult https://titleix.nd.edu/support-resources.

Accommodation for Disabilities

It is the policy and practice of The University of Notre Dame to provide reasonable accommodations for students with properly documented disabilities. Students who have questions about Sara Bea Accessibility Services or who have, or think they may have, a disability are invited to contact Sara Bea Accessibility Services for a confidential discussion by emailing at sarabeacenter@nd.edu or by phone at 574-631-7157. Because the University's Academic Accommodations Processes generally require students to request accommodations well in advance of the dates when they are needed, students who believe they may need an accommodation for this course are encouraged to contact Sara Bea Accessibility Services at their earliest opportunity. Additional information about Sara Bea Accessibility Services and to learn more about the student process for requesting accommodations, please visits Accessibility Support.

Mental Health Support

Diminished mental health can interfere with optimal academic performance. The source of symptoms might be related to your course work; if so, please speak with me. However, problems with other parts of your life can also contribute to decreased academic performance. The University Counseling Center (UCC) provides cost-free and confidential mental health services to help you manage personal challenges that threaten your emotional or academic well-being.

Remember, getting help is a smart and courageous thing to do — for yourself and for those who care about you. For more resources, please see ucc.nd.edu or studenthealth.nd.edu.

The UCC is located on the third floor of Saint Liam Hall
Phone: 574-631-7336. Hours: Monday-Friday 8:00am – 5:00pm. Urgent Crisis Line 24/7

Care and Wellness Consultants

Care and Wellness Consultants provide support and resources to students who are experiencing stressful or difficult situations that may be interfering with academic progress. Through Care and Wellness Consultants, students can be referred to The University Counseling Center (for cost-free and confidential psychological and psychiatric services from licensed professionals), University Health Services (which provides primary care, psychiatric services, case management, and a pharmacy), and The McDonald Center for Student Well Being (for problems with sleep, stress, and substance use). Visit supportandcare.nd.edu