

IE 490 – Introduction to Machine Learning and its Applications

Good to know

- **CRN:** 24229
- **Credits hours:** 3
- **Class days:** TR
- **Class times:** 3:00 P.M. – 4:15 P.M.
- **Location:** ARMS 1109
- **Prerequisites:** Calculus, Linear Algebra, Probability & Statistics, Programming

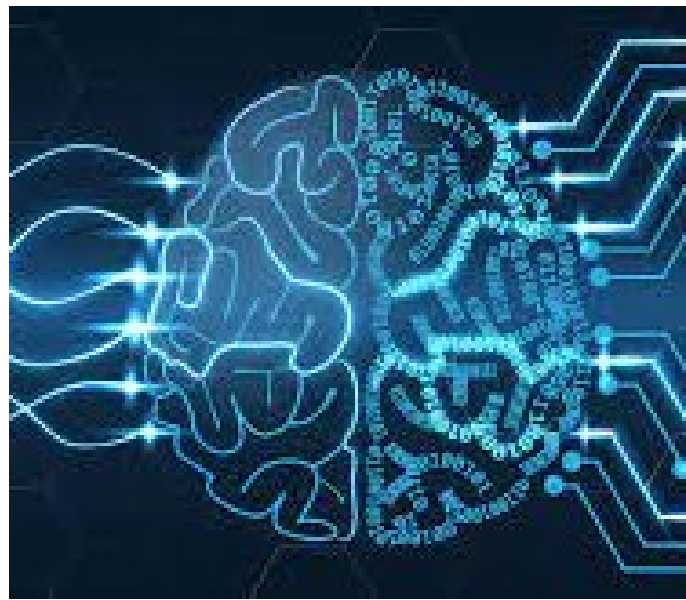
Course description

“Data scientist is the **sexiest job** of the 21st century” is the title of a 2012 Harvard Business Review article. What do data scientists do? They “**make discoveries** while swimming in data. It’s their preferred method of navigating the world. They are able to **bring structure** to large quantities of formless data and **make analysis** possible. In a competitive landscape where challenges keep changing and data never stop flowing, data scientists **help decision makers** shift from ad-hoc analysis to an ongoing **conversation with data.**”

In this course, you will be trained to swim in the growing ocean of data by using **machine learning** tools.

In **supervised learning**, the goal will be to predict the value of an outcome measure based on a number of input measures. For example, what is the estimated time a patient waits for an organ transplant, given his medical condition? In **unsupervised learning**, there is no outcome measure, and the goal is to describe the association and patterns among a set of input measures. For example, how many types of customers does a company have?

Data problems arise in many fields, such as: medicine, finance, and marketing. We have an exciting journey ahead! **Let’s dive in!**



Who is teaching?

- **Instructor:** Ana María Estrada Gómez
- **Office:** GRIS 364
- **Email:** amestrad@purdue.edu

Learning objectives

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

1. **Classify** problems as supervised or unsupervised learning.
2. **Classify** problems as regression or classification problems.
3. **Describe** different machine learning methods to solve these problems.
4. **Translate** a real-life problem into a machine learning one, using mathematical notation.
5. **Implement** the problem into a coding software (R, Python,...).
6. **Summarize** the results and conclude: what did the data teach us?
7. **Solve** a case study based on a realistic situation by **selecting** one of the different statistical models learned in class.

Evaluation

Graded Assignments	Description	Weight/Points
Attendance & participation	Learning sticks when it is effortful. Therefore, you are expected to participate fully in class activities.	4% (2% attendance, 2% participation)
Quizzes (8)	After each main topic, a quiz will be performed to assess if you are understanding ideas and concepts.	24% (3% each)
Homework (8)	Homework will include theory and coding problems. It will help you get familiarized with the methods presented in class. 'Real life' problems would be tackled. You will need to translate the problem into mathematical notation, implement it in a statistical software and summarize your findings.	48% (6% each)
Project	You can choose a case study relevant to you. You will need to select one or more of the statistical methods learned in class to propose a solution to the problem.	24%

Quizzes will be done at the beginning of class or on Brightspace.

The homework and project should be submitted through Brightspace, by the due date (see course schedule).

Grading scale

Letter grade	Percentage
A+	95.00 – 100.00
A	90.00 – 94.99
B+	85.00 – 89.99
B	80.00 – 84.99
C	70.00 – 79.99
D	60.00 – 69.99
F	< 60.00

Need help?

1. Brightspace discussion board

2. Hybrid office hour

- When: TBD
- Where: Option 1: Zoom (The link is in Brightspace). Option 2: TBD

3. In-person office hour

- When: Thursdays 4:30 P.M. – 5:30 P.M.
- Where: GRIS 364

4. Email

- Emails are read 9 a.m. to 5 p.m. daily and will be responded to within 24 hours.
- The subject line of your email should start with **IE 490**.

Materials for success

Text – Free online version

- James G., Witten D., Hastie T., Tibshirani R. (2021). An Introduction to Statistical Learning with applications in R (2nd edition). New York: Springer.

Additional references – Free online versions

- James G., Witten D., Hastie T., Tibshirani R., Taylor J. (2023). An Introduction to Statistical Learning with applications in Python. New York: Springer.
- Hastie T., Tibshirani R., Friedman J. (2009). The elements of statistical learning (2nd edition). New York: Springer.
- Bishop C.M. (2006). Pattern recognition and Machine Learning. New York: Springer.

Computer

In this course, you will solve data related problems using statistical software. We will apply the concepts learned through practical, in class, activities. The homework and project rely heavily on the use of a computer. If you don't own a laptop, please send the instructor an email as soon as possible to discuss possible solutions.

Software

You can choose to work with any of the following statistical software:

- R – free
- Python – free
- Matlab – free for Purdue students

In class, we will use R.

Course website

All relevant course information will be posted on Brightspace. Assignments should be submitted using Brightspace.

Guidelines

Course Expectations

Learning sticks when it is effortful. You elected to enroll in this class because you were driven to learn about machine learning. As a student this means you are expected to participate fully in class activities and make it a priority to complete readings and assignments for this class. As the instructor, I am expected to bring interesting opportunities for you to acquire the abilities of a data scientist. We will accomplish this in a positive, safe and interactive environment.

Attendance & Participation

Attendance is required in this course. Class participation is a crucial part of the learning process. I encourage you to be an active student in the classroom. Class participation and discussions are central to your own success and that of your colleagues. For these reasons, your attendance and participation account for 4% of your final grade.

Guidelines (continued)

Tardiness

It is extremely disruptive to the class when someone arrives late. Showing up late shows disrespect toward the instructor as well as your fellow students. Please be considerate!

Mobile devices

Research shows that unexpected noises and movement automatically divert and capture people's attention. This means that, by using your mobile device (laptop, cell phone, tablet, etc.), you could distract others, and disrupt their ability to learn. In addition, by using mobile devices you often become engaged in matters that are not related to the class. Further, research indicates that students taking notes on laptops tend to process less as they take notes, and the depth of their learning suffers.

Taking into account these facts, mobile devices are not allowed in the classroom, unless stated otherwise by the instructor. Your computer might be used only for class activities.

Approved absences: medically excused absence, grief, bereavement, military service, jury duty, and parenting leave.

Missed quizzes

No make-up quizzes will be allowed. If for any reason you miss ONE quiz, the grade will be the average of the remaining 7 quizzes. If you miss MORE than one quiz, the grade will be 0, unless it is an approved absence.

Late assignments

Late submission of homework is accepted at a 25% markdown, per day late, up to three days. No late assignment will be accepted after the solutions are posted online, with the exception of approved absences.

Collaboration & group work

We want you to help each other succeed. However, cheating and plagiarism, are expressly forbidden in this class. Any violation of course policies as it relates to academic integrity will result in a zero grade for that particular assignment and may result in a failing grade for the course. You will also be reported to the Office of Student Rights and Responsibilities.

Here are the things you need to know to ensure that you are working within the constraints of this course:

1. **Quizzes:** You may NOT use help from others (including AI agents).
2. **Homework:** You are welcome to work with others (including AI agents), but you must: list the names of anyone you worked with and write up your own solutions (including codes), based on your own understanding of the material.
3. **Project:** The project should be crafted and written by your group. You may talk with others about your ideas, but you should be working only with your team members on the actual outline and content. In addition, you are required to cite all sources you use.

Academic integrity

Purdue University aims to cultivate a community based on trust, academic integrity, and honor. You are expected to act according to the highest ethical standards. A hyperlink to Purdue's guide for academic integrity is included in Brightspace under University Policies.

Special accommodations

Purdue University is committed to making learning experiences accessible. If you experience barriers based on disability, you are welcome to let me know so that we can discuss options. You are also encouraged to contact the Disability Resource Center at drc@purdue.edu.

Guidelines (continued)

Nondiscrimination

In pursuit of its goal of academic excellence, the University seeks to develop and nurture diversity. The University believes that diversity strengthens the institution, stimulates creativity, promotes the exchange of ideas, and enriches campus life. A hyperlink to Purdue's full Nondiscrimination Policy Statement is included in our course Brightspace under University Policies.

Mental Health/Wellness

- If you find yourself beginning to feel some stress, anxiety and/or feeling slightly overwhelmed, try WellTrack
- If you find yourself struggling to find a healthy balance between academics, social life, stress, etc. sign up for free for one-on-one virtual or in-person sessions with a Purdue Wellness Coach at RecWell.
- If you are struggling and need mental health services contact Counseling and Psychological Services (CAPS) at 765-494-6995 during and after hours, on weekends and holidays, or by going to the CAPS office on the second floor of the Purdue University Student Health Center (PUSH) during business hours.

Basic Needs Security

Any student who faces challenges securing their food or housing and believes this may affect their performance in the course is urged to contact the Dean of Students for support. There is no appointment needed and Student Support Services is available to serve students 8 a.m.-5 p.m. Monday through Friday.

Emergency Preparation

In the event of a major campus emergency, course requirements, deadlines, and grading percentages are subject to changes that may be necessitated by a revised semester calendar or other circumstances beyond my control. Therefore, this syllabus is subject to change. You will be notified of any changes as far in advance as possible via an announcement on Brightspace. Monitor your Purdue email daily for updates.

Copyright

See the University Policies and Statements section of Brightspace for guidance on Use of Copyrighted Materials. Effective learning environments provide opportunities for students to reflect, explore new ideas, post opinions openly, and have the freedom to change those opinions over time.

Students and instructors are the authors of the works they create in the learning environment. As authors, they own the copyright in their works subject only to the university's right to use those works for educational purposes. Students may not copy, reproduce, or post to any other outlet (e.g., YouTube, Facebook, or other open media sources or websites) any work in which they are not the sole or joint author or have not obtained the permission of the author(s).

Course schedule

Class preparation: we expect you to prepare each class in advance. The material will be posted on Brightspace.

Quizzes will be held on Tuesdays, at the beginning of class or in Brightspace.

Homework is due on Fridays, before 11:59 P.M.

Date	Topic	Quizzes	Homework	Final project
9-Jan	Introduction			
11-Jan	Introduction			
16-Jan	Linear regression			
18-Jan	Linear regression			
23-Jan	Linear regression			
25-Jan	Invited speaker ML in industry			Team Assignment
30-Jan	Classification	Linear regression		
1-Feb	Classification		HW1	Topic preferences
6-Feb	Classification			
8-Feb	Classification			Topic assignment
13-Feb	Resampling methods	Classification		
15-Feb	Resampling methods		HW2	
20-Feb	Model selection			
22-Feb	Model selection			Project Proposal
27-Feb	Invited speaker ML in industry	Resampling methods and model selection		
29-Feb	Unsupervised learning		HW3	
5-Mar	Unsupervised learning			
7-Mar	Unsupervised learning			
12-Mar	Spring Break			
14-Mar				
19-Mar	Regression beyond linearity	Unsupervised learning		
21-Mar	Regression beyond linearity		HW4	
26-Mar	Tree-based methods	Regression beyond linearity		
28-Mar	Tree-based methods		HW5	Project Update
2-Apr	Support vector machines	Tree-based methods		
4-Apr	Support vector machines		HW 6	Project Feedback
9-Apr	Support vector machines			
11-Apr	Invited speaker ML in industry			
16-Apr	Deep learning	Support vector machines		
18-Apr	Deep learning		HW7	
23-Apr	Final project presentations	Deep learning		Presentations
25-Apr	Final project presentations		HW8	Presentations
30-Apr	Final exams week			
2-May				Final Report

Project: swimming in data (short version)

Throughout this project, you will test your swimming skills, you will solve a real end-to-end data science project. Don't worry, you won't be alone in the process, you will be working with three teammates. The first step will be to select a project. To help you, I have assembled a list of recommended projects. You can also propose a different project if you wish. You can find interesting datasets to work with in the following websites:

- Kaggle: <https://www.kaggle.com/>
- UCI Machine Learning Repository: <https://archive.ics.uci.edu>

The sequence of stages in the project are:

Stage	Description	Date
1	Project teams	Week 3 (assigned by the instructor)
2	Project preferences	Week 4
3	Project assignment	Week 5
4	Project proposal	Week 7
5	Project update	Week 12
7	Project presentation	Week 16
8	Project report	Thursday of exams week

Project teams

By the 3rd week of the semester, you will be assigned a team.

Project preferences

I know you would love to work on all the proposed problems. However, for the course purpose, one project will be enough. To guarantee that all problems are tackled, I want you to send a list of your team's top three preferred projects through Brightspace, before the end of the 4th week of class.

Project assignment

By the 5th week of the semester, I will post the project that your team has been assigned. I will try to make everyone happy. After this you are good to go. Happy swimming!

Project proposal (Graded: 20%)

By the 7th week of the semester, you will need to submit a document presenting the problem that you want to tackle, the machine learning methods that you plan to use, and the solution that you plan to obtain.

Project update 1 (Graded: 20%)

This is where the fun starts! By the 12th week of the semester, you will need to submit a project update presenting preliminary results, so that we all know where you are going, and all the possible obstacles that you may encounter in your journey.

Project presentation (Graded: 20%)

I want your classmates to hear about your journey swimming in data. Please prepare a presentation which summarizes your group's project and outcomes for the last week of classes.

Project report (Graded: 40%)

I hope you have enjoyed swimming in your data and making discoveries! Please submit your traveler's diary (project report) by Thursday of exams week.