Resources

DATA 322: Machine Learning for Data Science

Course Syllabus

DATA 322: Machine Learning for Data Science

Cal Poly Humboldt — Spring 2026

Course Information

Course Number: DATA 322
Course Title: Machine Learning for Data Science
Units: 4
Semester: Spring 2026
Mode of Instruction: In-person
Program Requirement: Required for the Data Science major

Lecture

Section: DATA 322-10 (CRN 22725)
Days/Time: Monday, Wednesday, Friday, 1:00–1:50 PM
Location: BSS 408

Laboratory (Activity)

Section: DATA 322-11 (CRN 22726)
Day/Time: Tuesday, 1:00–2:50 PM
Location: BSS 313

Final Exam

Final Exam: Wednesday, May 13, 2026, 12:40–2:30 PM
Location: BSS 408

Please check Canvas and your Cal Poly Humboldt email regularly for course announcements and updates.

Important Dates

Date	Event
January 20	Holiday (No class)
February 2	Add/Drop Deadline without a serious and compelling reason
February 20	Demo Mini-Project from 11-11:50 in BSS 166
March 8	Unsupervised Learning Project due
March 13	Midterm 1
March 16-20	Spring Break (No classes)
March 31	Cesar Chavez Day (No class)
May 10	Supervised Learning Project due

Date	Event
May 13	Final Exam: Wednesday, May 13, 2026, 12:40–2:30 PM, BSS 408

Note: This schedule is subject to change. Any changes will be communicated through **Canvas announcements**.

Instructor Information

Instructor: Dr. Rosanna Overholser
Email: rho3@humboldt.edu
Office Location: BSS 334
Office Phone: 707-826-4022

Office Hours:

Mondays 2:00–3:00 PM
Wednesdays 12:00–12:50 PM
Fridays 10:00–10:50 AM
Or by appointment

Please use your Cal Poly Humboldt email account to contact me.

Course Description (Catalog Description)

This course provides a broad introduction to machine learning, data mining, and statistical pattern recognition. Topics include supervised learning, unsupervised learning, and best practices in machine learning. The course draws from numerous case studies and applications, with a practical rather than theoretical emphasis.

Weekly: Lecture 3 hrs. Activity 2 hrs.
Prerequisites: DATA 271, MATH 107, STAT 109

A minimum grade of C- is required for the course to count toward the Data Science major.

Student Learning Outcomes

(Note, students can find program learning outcomes on the program website, and institution learning outcomes here: https://academicprograms.humboldt.edu/content/baccalaureate-student-learning-outcomes)

(1) Students will demonstrate computational skills to extract different types and quantities of data from multiple sources and create visualizations and other data products for various audiences. (ILO 3,4,7)

In order to achieve PLO 1, students will demonstrate the following specific measurable behaviors:

- (SLO 1) Students will extract quantitative and categorical data from a large public database, calculate relevant summary statistics, and present findings numerically and graphically for in-class individual and group data projects.

(2) Students will demonstrate statistical knowledge to build mathematical models and ensure the validity of data and its analysis. (ILO 7)

In order to achieve PLO 2, students will demonstrate the following specific measurable behaviors:

- (SLO 2) Students will design a mathematical (logic) model to guide analysis and ensure valid and reliable results for in-class individual and group data projects.

(5) Students will communicate effectively with a diverse range of audiences. (ILO 5,6)

In order to achieve PLO 5, students will demonstrate the following specific measurable behaviors:

- (SLO 5a) Students will prepare a written description of a mathematical model in which they explain the context, describe their methodology, analyze their data, and reach a conclusion that includes a discussion of its implications.

Additionally, one of students' two projects, communicated in written form, will be the application of Machine Learning algorithms (supervised and unsupervised, respectively) to a data set related to one of the following: sustainability, social justice, or Traditional Ecological Knowledge (TEK). This project supports ILOs (1) and (2).

Course Learning Outcomes

By the end of this course, students will be able to:

1. Recognize the characteristics of machine learning that make it useful for real-world problems.
2. Explain how machine learning differs from inferential statistical modeling.
3. Characterize machine learning algorithms as supervised and unsupervised.
4. Implement common machine learning algorithms using Python-based toolkits.
5. Apply regression, classification, and clustering methods to real data sets.
6. Evaluate model performance and select appropriate models based on context and goals.
7. Communicate modeling choices, results, and limitations to a variety of audiences.
8. Identify and discuss ethical considerations related to data collection, modeling, and algorithmic decision-making.

Program & Institutional Learning Outcomes

This course supports the Data Science Program Learning Outcomes and Cal Poly Humboldt's Institutional Learning Outcomes related to quantitative reasoning, critical thinking, communication, and ethical reasoning.

Course Materials & Workload

DATA 322 includes:

- Three weekly lecture meetings
- One weekly laboratory session

In this 4-unit course, students should expect to spend approximately **12 hours per week** on course-related work. This includes time in lecture (3 hours), lab time (2 hours), and time outside of class to complete readings, assignments, projects, and exam preparation (approximately 7 hours).

Required Materials

Textbook:

An Introduction to Statistical Learning with Applications in Python

Available free online or for purchase in print.

Software:

Python and commonly used data science libraries (installed on all campus computers)

Other Technology:

Regular access to a computer with reliable internet

Students without personal access to required technology may use campus computer labs and library resources.

How to Succeed in This Course

Students enter DATA 322 with different backgrounds in statistics, mathematics, and programming. That's expected. You do not need to be an expert at the start of the semester to succeed in this course. What matters most is steady engagement and a willingness to practice, reflect, and ask questions.

Students who are successful in this course typically:

- **Attend lecture and lab regularly.**
  Class time is used to build intuition, practice ideas, and work through examples that are difficult to learn independently.
- **Participate actively in daily warm-ups.**
  Warm-ups are designed to help you engage with your classmates and develop conceptual understanding and confidence before moving to more complex problems.
- **Use labs as a space to practice and make mistakes.**
  Labs are meant for learning, not perfection. Comparing your work to posted results and reflecting on differences is an important part of the learning process.
- **Prepare consistently for weekly quizzes.**
  Quizzes are short and predictable, but they require regular practice. Reviewing lecture notes, lab activities, and assigned textbook problems each week will help you succeed.
- **Start projects early and work iteratively.**
  Projects are most successful when you begin early, test ideas incrementally, and seek feedback along the way. I've deliberately chosen two projects that reflect significant problems with real applications of ML. At first glance, the datasets might appear simple but great care is required to build useful products out of them. Rather than tell you a solution to implement, I'm going to use our class time to guide class discussions so that students will be active participants in creating appropriate solutions.
- **Collaborate appropriately with classmates.**
  Discussing ideas and approaches is encouraged, but all submitted work must reflect your own understanding unless otherwise specified.
- **Ask for help when you need it.**
  If you are struggling with course material, please reach out during office hours, lab sessions, or by email. Seeking help early can make a significant difference.

As a general guideline, you should plan to spend approximately **three hours per week per unit** on course-related work, including time in class and time outside of class. For this 4-unit course, this corresponds to about **12 hours per week** total. Individual experiences will vary, but this can be a helpful frame of reference when planning your time.

Course Feedback & Assessment

Students will complete a variety of assessments designed to support regular practice, provide regular feedback, and evaluate mastery of machine learning concepts and methods.

Assessments in this course include:

- **Daily warm-up problems** during lecture (Mondays, Wednesdays, and Fridays).
  These short activities are designed to help students engage with new concepts, build intuition, and practice explaining ideas with peers. Warm-ups are graded on participation.
- **Weekly programming labs**.
  Labs provide guided, hands-on practice implementing machine learning methods using Python. Labs are intended

as formative learning experiences. I'll provide desired outputs so students can check their work and reflect on their understanding before turning in their work. Labs will be graded on completion.

Early in the semester, you will complete a **mini project** as part of the weekly labs. This project introduces the full workflow of an applied machine learning task in a simplified and supported setting. Using a provided Google Colab notebook, the class will modify an existing recommender system to form study groups based on survey responses collected from real participants.

The goal of this mini project is not to produce perfect groupings, but to practice:

  - Designing questions that can be translated into quantitative features
  - Explore distance metrics
  - Applying and interpreting clustering algorithms
  - Working collaboratively with a shared product
  - Reflecting on modeling decisions and their consequences

This experience is designed to make the larger course projects more approachable and to build confidence before tackling more complex open-ended problems.

- **Weekly quizzes** during lab sessions.
  Quizzes assess understanding of both lab material and assigned practice problems. Quiz questions may include conceptual questions, mathematical calculations, interpretation of output or visualizations, and short problem-solving tasks. Quizzes are designed to encourage steady preparation and synthesis of ideas.
- **Two projects**.
  Projects require students to apply machine learning methods to real data, justify modeling choices, evaluate performance, and communicate results in written form. Projects emphasize modeling choices, interpretation, and communication and may be used as part of a professional portfolio.
  - **Unsupervised Learning Project:**
    In this project, you will use unsupervised learning techniques to build a *major recommender system* for Cal Poly Humboldt using publicly available data. The goal is to explore and justify modeling choices, and communicate the resulting recommendations clearly and responsibly.
  - **Supervised Learning Project:**
    In this project, you will work with data from a one-dimensional simulation of a fusion reactor to develop a model to predict the onset of instabilities *before* they occur. This project emphasizes feature engineering, computational efficiency and valid model evaluation in an effort to help further research in California's renewable energy sector.
- **One midterm exam and one final exam**.
  Exams assess students' ability to reason about machine learning methods, interpret results, compare approaches, and understand limitations. Exams focus on conceptual understanding rather than memorization of code syntax.

Grading

Component	Weight
Daily Participation (Warm-ups)	5%
Weekly Quizzes	25%
Weekly Labs	15%
Two Projects	30%
Midterm Exam	10%
Final Exam	15%
Total	**100%**

Final letter grades will be assigned according to the following scale:

A: 93–100% | A-: 90–92% | B+: 87–89% | B: 83–86% | B-: 80–82% | C+: 77–79% | C: 73–76% | C-: 70–72% | D:

60–69% | F: 0–59%

Grades are recorded in Canvas. Canvas grades are not official until finalized at the end of the semester.

Attendance & Participation

Regular attendance and active participation are expected and essential for success in this course. Attendance in lecture and lab will be reflected in participation activities and assignments.

If you must miss class, you are responsible for obtaining notes and keeping up with course material.

Late or Missed Work Policies

- **Daily warm-ups** are graded on participation and cannot be made up if missed.
- **Weekly labs** are due on the posted deadlines in Canvas. Late submissions may be accepted in documented emergencies; otherwise, late work policies will be specified per lab.
- **Weekly quizzes** are administered during lab sessions. If you need to miss a quiz due to illness or other emergencies, please contact me to make an arrangement to take it at another time.
- **Projects** are due on the specified dates in Canvas. Late project policies will be specified per project.
- **Exams** must be taken on the scheduled date unless arrangements are made in advance due to illness or emergency.

Technology issues are not automatically considered valid excuses for late or missed work.

Course Topics (Approximate Order)

- Data preprocessing skills from DATA 271
- Introduction to machine learning
- Unsupervised learning
- Traditional supervised learning methods
- Model evaluation and validation
- Modern supervised learning methods
- Feature engineering & computational efficiency
- Ethical considerations and appropriate communication of results

---

University Policies & Resources

Students are responsible for reviewing the **Cal Poly Humboldt Syllabus Addendum**, which includes policies and resources related to:

- Academic honesty
- Disability accommodations
- Emergency procedures
- Title IX
- Student conduct and support services

The Syllabus Addendum is available on the Cal Poly Humboldt website.

Accommodations

Students who wish to request disability-related accommodations should contact the Student Disability Resource Center (SDRC) as early as possible and then email me your letter of accommodation.

Classroom & Online Conduct

Standards of respectful and professional behavior apply to all course interactions, including in-person classes, assignments, and email communication.

Changes to the Syllabus

This syllabus is subject to change. Any changes will be communicated through Canvas announcements.

Course Schedule

Spring 2026

Note: This schedule is subject to change. Any changes will be communicated through **Canvas announcements**.

Week	Mon	Tue	Wed	Fri
1 (Jan 20–23)	Holiday	Lab 1	Lec	Lec
2 (Jan 26–30)	Lec	Lab 2 + Quiz 1	Lec	Lec
3 (Feb 2–6)	Lec	Lab 3 + Quiz 2	Lec	Lec
4 (Feb 9–13)	Lec	Lab 4 + Quiz 3	Lec	Lec
5 (Feb 16–20)	Lec	Lab 5 + Quiz 4	Lec	Demo Mini-Project (11-11:50, BSS 166)
6 (Feb 23–27)	Lec	Lab 6 + Quiz 5	Lec	Lec
7 (Mar 2–6)	Lec	Lab 7 + Quiz 6	Lec	LecProject 1 due Sun, Mar 8
8 (Mar 9–13)	Lec	Quiz 7	Lec	Midterm 1
— (Mar 16–20)	—	—	—	—
9 (Mar 23–27)	Lec	Lab 8 + Quiz 8	Lec	Lec
10 (Mar 30–Apr 3)	Lec	Holiday (Cesar Chavez Day)	Lec	Lec
11 (Apr 6–10)	Lec	Lab 9 + Quiz 9	Lec	Lec
12 (Apr 13–17)	Lec	Lab 10 + Quiz 10	Lec	Lec
13 (Apr 20–24)	Lec	Lab 11 + Quiz 11	Lec	Lec
14 (Apr 27–May 1)	Lec	Lab 12 + Quiz 12	Lec	Lec
15 (May 4–8)	Lec	Quiz 13	Lec	LecProject 2 due Sun, May 10
Finals (May 11–15)	—	—	Final Exam	—

Important Dates:

- **January 20 -** Holiday (No class)
- **February 2 -** Add/Drop Deadline without a serious and compelling reason
- **February 20 -** Demo Mini-Project (11-11:50 AM, BSS 166)
- **March 8 -** Unsupervised Learning Project due
- **March 13 -** Midterm 1
- **March 16-20 -** Spring Break (No classes)
- **March 31 -** Cesar Chavez Day (No class on Tuesday)
- **May 10 -** Supervised Learning Project due
- **May 13 -** Final Exam (12:40–2:30 PM, BSS 408)

Assessments

Overview

This course includes several types of assessments designed to support regular practice, provide regular feedback, and evaluate mastery of machine learning concepts and methods.

Component	Weight	Description
Daily Participation (Warm-ups)	5%	Warm-up problems during lecture (Mondays, Wednesdays, Fridays)
Weekly Quizzes	25%	Quizzes during lab sessions (Tuesdays)
Weekly Labs	15%	Programming labs implementing machine learning methods using Python
Two Projects	30%	Unsupervised Learning Project and Supervised Learning Project
Midterm Exam	10%	Midterm 1 (March 13)
Final Exam	15%	Final Exam (May 13, 12:40–2:30 PM, BSS 408)
Total	**100%**	

Daily Participation (Warm-ups) (5%)

Daily warm-up problems are completed during lecture (Mondays, Wednesdays, and Fridays). These short activities are designed to help students engage with new concepts, build intuition, and practice explaining ideas with peers. Warm-ups are graded on participation.

Weekly Labs (15%)

Weekly programming labs provide guided, hands-on practice implementing machine learning methods using Python. Labs are intended as formative learning experiences. I'll provide desired outputs so students can check their work and reflect on their understanding before turning in their work. Labs will be graded on completion.

Early in the semester, you will complete a **mini project** as part of the weekly labs. This project introduces the full workflow of an applied machine learning task in a simplified and supported setting. Using a provided Google Colab notebook, the class will modify an existing recommender system to form study groups based on survey responses collected from real participants.

Individual Labs

- Lab 1 - Instructions coming soon
- Lab 2 - Instructions coming soon
- Lab 3 - Instructions coming soon
- Lab 4 - Instructions coming soon

- Lab 5 - Instructions coming soon
- Lab 6 - Instructions coming soon
- Lab 7 - Instructions coming soon
- Lab 8 - Instructions coming soon
- Lab 9 - Instructions coming soon
- Lab 10 - Instructions coming soon
- Lab 11 - Instructions coming soon
- Lab 12 - Instructions coming soon

Weekly Quizzes (25%)

Weekly quizzes are administered during **lab sessions (Tuesdays)**. Quizzes assess understanding of both lab material and assigned practice problems. Quiz questions may include conceptual questions, mathematical calculations, interpretation of output or visualizations, and short problem-solving tasks. Quizzes are designed to encourage steady preparation and synthesis of ideas.

If you need to miss a quiz due to illness or other emergencies, please contact me to make an arrangement to take it at another time.

Individual Quizzes

- Quiz 1 - Instructions coming soon
- Quiz 2 - Instructions coming soon
- Quiz 3 - Instructions coming soon
- Quiz 4 - Instructions coming soon
- Quiz 5 - Instructions coming soon
- Quiz 6 - Instructions coming soon
- Quiz 7 - Instructions coming soon
- Quiz 8 - Instructions coming soon
- Quiz 9 - Instructions coming soon
- Quiz 10 - Instructions coming soon
- Quiz 11 - Instructions coming soon
- Quiz 12 - Instructions coming soon
- Quiz 13 - Instructions coming soon

Projects (30%)

Two projects throughout the semester require students to apply machine learning methods to real data, justify modeling choices, evaluate performance, and communicate results in written form. Projects emphasize modeling choices, interpretation, and communication and may be used as part of a professional portfolio.

Late project policies will be specified per project.

Individual Projects

- **Unsupervised Learning Project** - Due March 8
  In this project, you will use unsupervised learning techniques to build a *major recommender system* for Cal Poly Humboldt using publicly available data. The goal is to explore and justify modeling choices, and communicate the resulting recommendations clearly and responsibly.
- **Supervised Learning Project** - Due May 10
  In this project, you will work with data from a one-dimensional simulation of a fusion reactor to develop a model to predict the onset of instabilities *before* they occur. This project emphasizes feature engineering, computational efficiency and valid model evaluation in an effort to help further research in California's renewable energy sector.

Exams (25%)

Midterm Exam (10%)

- **Midterm 1:** March 13 (in lecture)

Exams assess students' ability to reason about machine learning methods, interpret results, compare approaches, and understand limitations. Exams focus on conceptual understanding rather than memorization of code syntax.

If you need to miss a midterm due to illness or other emergencies, please communicate with me as soon as possible to reschedule it.

Final Exam (15%)

Final Exam: Wednesday, May 13, 2026, 12:40–2:30 PM, BSS 408

The final exam is required and assesses comprehensive understanding of machine learning concepts and methods covered throughout the semester.

Grading Scale

Final letter grades will be assigned based on overall performance:

A: 93-100% | A-: 90-92% | B+: 87-89% | B: 83-86% | B-: 80-82% | C+: 77-79% | C: 73-76% | C-: 70-72% | D: 60-69% | F: 0-59%

A minimum grade of **C-** is required for the course to count toward the Data Science major.

Lectures

Week 1

Monday

Holiday - No class

Wednesday

Lecture 1

Friday

Lecture 2

Week 2

Monday

Lecture 3

Wednesday

Lecture 4

Friday

Lecture 5

Week 3

Monday

Lecture 6

Wednesday

Lecture 7

Friday

Lecture 8

Week 4

Monday

Lecture 9

Wednesday

Lecture 10

Friday

Lecture 11

Week 5

Monday

Lecture 12

Wednesday

Lecture 13

Friday

Lecture 14

Week 6

Monday

Lecture 15

Wednesday

Review

Friday

Midterm 1

Week 7

Monday

Lecture 16

Wednesday

Lecture 17

Friday

Lecture 18

Week 8

Monday

Lecture 19

Wednesday

Lecture 20

Friday

Lecture 21

Week 9

Monday

Lecture 22

Wednesday

Lecture 23

Friday

Lecture 24

Week 10

Monday

Lecture 25

Wednesday

Lecture 26

Friday

Lecture 27

Week 11

Monday

Lecture 28

Wednesday

Lecture 29

Friday

Lecture 30

Week 12

Monday

Lecture 31

Wednesday

Lecture 32

Friday

Lecture 33

Week 13

Monday

Lecture 34

Wednesday

Review

Friday

Midterm 2

Week 14

Monday

Lecture 35

Wednesday

Lecture 36

Friday

Lecture 37

Week 15

Monday

Lecture 38

Wednesday

Review

Friday

Review

Methods Map

Python Quick Reference

[Content to be added - Python commands and functions for machine learning]

Import Data

[Content to be added]

Summaries

[Content to be added]

Plots

[Content to be added]

Simulation

[Content to be added]

Resources

This page provides links to essential course resources including the textbook, Python software downloads, and reference materials.

Lectures

Lectures - Access all course lectures organized by week and day.

Textbook

Required Textbook: *An Introduction to Statistical Learning with Applications in Python*

Available free online or for purchase in print. This textbook provides a broad and less technical treatment of key topics in statistical learning, with practical applications in Python.

Python Software Downloads

DATA 322 uses Python for machine learning and data analysis. You can use Python in several ways:

Python (Base Installation)

Download Python: Python.org

Python is free and open-source. The base Python installation includes the core language and standard library.

- **Windows:** Download the installer from python.org
- **Mac:** Python may be pre-installed, or download from python.org
- **Linux:** Install via your distribution's package manager or from python.org

Python Data Science Libraries

Commonly used libraries for this course include: - `numpy` - Numerical computing - `pandas` - Data manipulation and analysis - `scikit-learn` - Machine learning algorithms - `matplotlib` - Data visualization - `seaborn` - Statistical data visualization

Install via pip: `pip install numpy pandas scikit-learn matplotlib seaborn`

Python in Google Colab

Access Python in Google Colab: Google Colab

Google Colab provides free access to Python in a cloud-based notebook environment with many data science libraries pre-installed. This is useful if you don't want to install Python on your computer or need to work from different devices.

Python in Jupyter Notebook

Install Jupyter: Jupyter Installation Guide

Jupyter Notebook is a popular environment for interactive Python development and data analysis.

1. Install Python first
2. Install Jupyter: `pip install jupyter`
3. Launch Jupyter: `jupyter notebook`

Note: Python and commonly used data science libraries are installed on all campus computers, so you can use the computer labs if you prefer not to install software on your personal computer.

Course Reference Materials

Python Reference

Python Quick Reference - A quick reference guide for common Python commands and functions used in this course for machine learning and data analysis.

Methods Map

Methods Map - A reference table of machine learning methods covered in the course, organized by topic.

Additional Resources

Campus Resources

- **Computer Labs:** Python and commonly used data science libraries are installed on all campus computers
- **ROSE (Reusable Office Supply Exchange):** Free used computers may be available if you need a computer

Getting Help

- **Office Hours:** See the Syllabus for instructor office hours
- **Canvas:** Check Canvas for announcements and course communications
- **Email:** Contact the instructor at rho3@humboldt.edu