



Machine Learning CS 4641 & 7641

Spring 2027

PACKET SUMMARY

26 selected pages included in this export.

Schedules are filtered automatically for the active instructors in the current semester.

EXPORT NOTES

Use your browser's Print dialog and choose Save as PDF for a clean letter-sized packet.

Turn off Headers and footers. Turn on Background graphics if your browser offers it.

INCLUDED PAGES

COURSE INFO

3 pages selected

Course Info • Course Overview • Course Schedule (Max Mahdi)

GUIDELINES

6 pages selected

Guidelines • General • Office Hours • Project Interview Rules • +2 more

GRADING

15 pages selected

Grading • Categories • Quizzes and Exams • Projects: Individual and Team •
+11 more

RESOURCES

2 pages selected

Resources • Collection

COURSE INFO



Course Info



Course Overview

Course Overview

Important

CSE students should note that CS 7641 is not allowed as a substitute for the CSE core course CSE 6740, and that they cannot get credit for both CSE 6740 and CS 7641.

Course Description

This course introduces techniques in machine learning with an emphasis on *algorithms and their applications to real-world data*. We will investigate the following question: how to computationally extract useful knowledge from data for decision making and task support! The course will also cover briefly Ethics in Machine Learning and Secure Computing. We will focus on machine learning methods, which are organized into three main course objectives:

1. **Basic math for data science and machine learning**
 - Linear algebra
 - Probability and statistics
 - Information theory
 - Optimization
2. **Unsupervised machine learning for data exploration**
 - Clustering analysis
 - Dimensionality reduction
 - Kernel density estimation
3. **Supervised learning for predictive data analysis**
 - Tree-based models
 - Support vector machines
 - Linear classification and regression
 - Neural networks

Prerequisites for this course include (1) basic knowledge of probability, statistics, and linear algebra; (2) basic programming experience in Python.

Course Objectives and Learning Outcomes

- Introduce to you the pipeline of Machine Learning
- Help you understand major machine learning algorithms
- Help you learn to apply tools for real data analysis problems
- Encourage you to do research in data science and machine learning

In addition to the technical content, this class includes the following learning objectives:

- Structuring a task into a machine learning work flow
- Collaborating effectively in a remote learning environment
- Conducting peer evaluation in a constructive format
- Communicating technical content in a concise and effective manner

COURSE INFO

Course Schedule (Max Mahdi)

Schedule

Important

All deadline and due dates in this course will be at 11:59 PM ET.

Scroll horizontally to see the full schedule table on mobile devices

Week	Dates	Topics (Available on canvas)	Quizzes	Exam	Individual & team projects	Readings
1	Jan 11th - Jan 15th	<ul style="list-style-type: none">Course introduction (L1)NumPy and broadcasting (L2)	<ul style="list-style-type: none">Q0: L1, L2 (warmup)		<ul style="list-style-type: none">Individual and team projects released: Jan 15th	<ul style="list-style-type: none">GT Honor CodeHeilmeier CatechismGitHub Pages QuickstartYAML ConfigurationNumPy Tutorialseaborn: statistical data visualizationOverleaf for GT studentsNumPy tutorial
2	Jan 18th - Jan 22nd	<ul style="list-style-type: none">Official institute holiday – Martin Luther King, Jr. Day (Jan 18th - Jan 18th)Linear algebra (L3)	<ul style="list-style-type: none">Q1: L3; Syllabus quiz			<ul style="list-style-type: none">Linear Algebra Review by Zico KolterStanford linear algebra review
3	Jan 25th - Jan 29th	<ul style="list-style-type: none">Probability and statistics (L4)	<ul style="list-style-type: none">Q2: L4		<ul style="list-style-type: none">Team composition due: Jan 31st	<ul style="list-style-type: none">Independence, Correlation, OrthogonalityStanford probability review
4	Feb 1st - Feb 5th	<ul style="list-style-type: none">Information theory (L5)Optimization (L6)*Jensen Inequality Interactive Visualization*Lagrange FunctionToolbox, part 2 (L7)	<ul style="list-style-type: none">Q3: L5, L6		<ul style="list-style-type: none">Team proposal due: Feb 7th	<ul style="list-style-type: none">The Differences Between Data, Information and KnowledgeVisual Information Theory by Chris OlahMore about Cross Entropy and KLDKKT for inequality constrained optimizationVisual information theoryStanford optimization notes
5	Feb 8th - Feb 12th	<ul style="list-style-type: none">Clustering and K-means (L8)	<ul style="list-style-type: none">Q4: L7, L8			<ul style="list-style-type: none">Matplotlib Tutorialsseaborn: statistical data visualizationClustering guide

Week	Dates	Topics (Available on canvas)	Quizzes	Exam	Individual & team projects	Readings
6	Feb 15th - Feb 19th	<ul style="list-style-type: none"> Gaussian mixture models (L9) 	<ul style="list-style-type: none"> Q5: L9 			<ul style="list-style-type: none"> Gaussian mixture models
7	Feb 22nd - Feb 26th	<ul style="list-style-type: none"> Hierarchical clustering (L10) Density-based clustering / DBSCAN (L11) 	<ul style="list-style-type: none"> Q6: L10, L11 			<ul style="list-style-type: none"> Dendrogram Visualization Clustering guide
8	Mar 1st - Mar 5th	<ul style="list-style-type: none"> Clustering evaluation (L12) Density estimation (L13) Dimensionality reduction (L14) 	<ul style="list-style-type: none"> Q7: L12 Q8: L13, L14 			<ul style="list-style-type: none"> Clustering evaluation KDE interactive visualization KernelDensity in scikit-learn Feature extraction using Principal Component Analysis Principal Component Analysis for images Density estimation PCA and decomposition
9	Mar 8th - Mar 12th	<ul style="list-style-type: none"> Spring Break (Mar 8th - Mar 12th) 				
10	Mar 15th - Mar 19th	<ul style="list-style-type: none"> Midterm review 		<ul style="list-style-type: none"> Midterm exam opens: through L14; Canvas has exam instructions: Mar 15th Midterm exam closes (11:59 PM ET): Mar 21st 	<ul style="list-style-type: none"> Team midterm review (update your website before your meeting): Mar 15th - Mar 28th 	
11	Mar 22nd - Mar 26th	<ul style="list-style-type: none"> Linear regression (L15) Regularized regression (L16) 	<ul style="list-style-type: none"> Q9: L15, L16 		<ul style="list-style-type: none"> Individual midterm: 10-minute unlisted YouTube video and website update due (11:59 PM ET): Mar 28th 	<ul style="list-style-type: none"> Adding Noise to Regression Predictors Linear models
12	Mar 29th - Apr 2nd	<ul style="list-style-type: none"> Logistic regression (L17) 	<ul style="list-style-type: none"> Q10: L17 			<ul style="list-style-type: none"> Why Cross Entropy over MSE for Classification More about Cross Entropy and KLD Linear models
13	Apr 5th - Apr 9th	<ul style="list-style-type: none"> Practical advice (L18) Neural networks (L19) 	<ul style="list-style-type: none"> Q11: L19 			<ul style="list-style-type: none"> Neural Network Playground Initializing neural networks Back propagation numerical example Common ML pitfalls Stanford neural networks notes

Week	Dates	Topics (Available on canvas)	Quizzes	Exam	Individual & team projects	Readings
14	Apr 12th - Apr 16th	<ul style="list-style-type: none"> Convolutional neural networks (L20) Decision trees (L21) Ensemble learning and random forests (L22) 	<ul style="list-style-type: none"> Q12: L20 Q13: L21, L22 		<ul style="list-style-type: none"> Individual final interview: meet with your TA (no individual final video): Apr 12th - Apr 25th 	<ul style="list-style-type: none"> Convolutional Neural Network Live Demo Back Propagation in Convolutional Neural Networks Stanford CNN notes Decision trees Ensemble methods
15	Apr 19th - Apr 23rd	<ul style="list-style-type: none"> Support vector machines (L23) SVM kernels (L24) Supervised learning review 	<ul style="list-style-type: none"> Q14: L23, L24 		<ul style="list-style-type: none"> Team final video due (one unlisted YouTube link per team; no interview; 11:59 PM ET): Apr 25th 	<ul style="list-style-type: none"> Support vector machines
16	Apr 26th - Apr 27th	<ul style="list-style-type: none"> Final instructional days: Apr 26th - Apr 27th Course review and project completion 			<ul style="list-style-type: none"> Individual and team final websites due (last instructional day, 11:59 PM ET): Apr 27th Project award consideration based on final websites: Apr 27th 	

GUIDELINES



Guidelines



General

Online participation and course materials

Core lectures are pre-recorded; live lecture attendance is not required. Study the assigned videos and notes, complete assessments within their availability windows, and work on both the individual and team projects. The schedule also links interactive visualizations and demonstrations when available.

Course-created videos, notes, assessment questions, and code are for enrolled students. Do not redistribute them outside the course without written permission.

Course platforms

- **Canvas** hosts lecture modules, quizzes, the midterm exam, and the gradebook. Canvas timestamps control Canvas assessment submissions.
- **ML Class Hub** manages project teams, project links and submissions, accommodation-document intake, and the [midterm reference-sheet PDF](#).
- **Ed Discussion** is the official announcement and course-question channel. Its page explains notifications, public and private posts, and assessment-discussion limits.

Deadlines and time zones

Website deadlines are 11:59 PM Eastern Time (ET), Atlanta local time, unless stated otherwise. ET changes between EST and EDT. Canvas can display deadlines in your account's time zone; the exact Canvas deadline controls Canvas assessments. Check and convert times before an availability window closes. The course [schedule](#) lists weekly dates.

Academic integrity and generative AI

The [Georgia Tech Academic Honor Code](#) applies to all work. Quizzes and the midterm must be completed independently without generative AI or unauthorized assistance; see [Quizzes and Exams](#) for permitted materials.

Collaboration within a registered team and disclosed use of generative AI are allowed for the projects under the [project rules](#). The individual project must represent the student's own understanding.

For either project, disclose each AI tool, its purpose, and the scope of its contribution. Verify generated code, text, data, results, and references. Do not put sensitive, protected, regulated, confidential, licensed, or personally identifiable information into an AI system. Every citation must identify a real source that you reviewed and that supports the cited claim. Do not fabricate sources or results. Suspected violations are referred to the Georgia Tech Office of Student Integrity; that office determines responsibility under Institute procedures.

Accommodations and exceptional circumstances

Disability accommodations require approval through the [Office of Disability Services](#) and an official Faculty Notification Letter. Request that the letter be sent through the Institute process, then submit the ODS request and PDF in [ML Class Hub Accommodations](#) for course-team verification. A student-uploaded copy alone does not grant approval.

For an exceptional circumstance or assessment make-up, first use the [Dean of Students Request Assistance process](#). Upload the official notification as a PDF in ML Class Hub Accommodations; an official email may be saved or printed to PDF. Then notify your assigned mentor through a private Ed post identifying the affected requirement; follow the [contact instructions](#). Do not send medical records or diagnoses to course staff or Ed. An informal message alone does not approve a make-up. Contact the Institute and course team promptly; a letter received more than one week after the affected deadline cannot support a retroactive course exception under the current policy.

Regrades and changes

Follow the instructions and deadline published with an assessment for a specific grading correction. A full regrade can raise, leave unchanged, or lower the score. Course-wide changes are announced on Ed and reflected in the relevant course page or Canvas item.



Office Hours

Online Office Hours

Office hours are optional online help sessions. They do not replace scheduled project reviews.

Joining a session

1. A TA announces the session in an Ed thread with the time and any focus topic.
2. Reply with a short, specific question and what you have tried. Do not post private information, complete code, grades, or confidential project data.
3. The TA answers in the thread when useful to the class, or sends a private meeting link when screen sharing or individual discussion is needed.
4. The TA works through the queue as time allows and posts when the session ends. Joining the queue does not guarantee service before the session ends.

Come prepared with the relevant lecture, error, small code example, or project requirement. TAs can explain concepts, isolated errors, and requirements; they do not debug an entire repository, design a project for a team, provide active assessment answers, or pre-grade a submission. See [Ed Discussion](#) for posting and privacy rules.



Project Interview Rules

Project Interview Rules

These rules apply to **both the team midterm interview and the individual final interview**.

- **Camera on:** every participating student must keep their camera on with their face clearly visible throughout the interview, including all members attending a team interview. If an approved accommodation or a technical problem affects this requirement, contact your assigned mentor through a private Ed Discussion post before the interview when possible (see [contact instructions](#)). Report an unexpected problem to the interviewing TA immediately and follow up privately on Ed.
- **No live AI assistance:** do not interact with ChatGPT, Claude, Gemini, Copilot, or any other generative-AI assistant during the interview. This includes browser tabs, phone apps, voice assistants, coding assistants, and tools that listen to questions or generate suggested answers. Close or disable them before joining. Permission to use and disclose AI while developing a project does not permit AI assistance during its interview.
- **Answer from your own understanding:** do not obtain answers through search engines, messaging platforms, another person, or a second device. You may show your project website, code, and results as directed by the TA. In a team interview, each member must answer questions directed to them; teammates must not privately send answers or speak for that member. Team discussion is allowed when the TA invites it.
- **Be ready to demonstrate your work:** the TA may ask you to share the relevant project window, explain code or results, and make a small change live. Close unrelated private content before sharing. Do not share passwords or private documents.
- **Interview observation:** the TA observes participation, responses, and activity visible on camera or the shared screen and may ask you to explain an apparent off-screen interaction. This is human observation, not automated eye tracking. Looking away or moving your eyes alone is not

proof of unauthorized assistance. Any concern will be reviewed in context, with an opportunity for the student to explain, under the course's academic-integrity process.



Ed Discussion

Ed Discussion

Ed is the official channel for course announcements, class-wide clarifications, questions, and office-hour notices. Join through Canvas, enable notifications, and read pinned posts and megathreads before posting. Use a descriptive title and say which lecture, assessment, or project requirement your question concerns.

Contact your mentor through Ed Discussion

Use Ed Discussion for all course questions and communication with teaching staff. Contact your assigned mentor first. Identify your mentor in your Ed post so it can be routed appropriately. Your mentor will involve another TA or the instructor when needed. If you do not know your assigned mentor, ask the teaching staff in a private Ed post.

Use a private Ed post for personal or sensitive matters. Do not use email, Canvas messages, or ML Class Hub messages to request course support. The Hub remains the place for required submissions and official accommodation documents; use Ed for the accompanying communication. For a problem during a proctored assessment, follow the [Honorlock incident procedure](#) first and contact your mentor on Ed afterward.

Public and private posts

Use a public post when the answer can help classmates, such as a conceptual question, a broken link, or a project-requirement clarification. Use a private post for grades, accommodations, emergencies, team conflicts, technical incidents, or possible misconduct. Do not include medical documents, personal data, complete solutions, or private grades in a public post.

Ask a focused question and explain what you tried. The team can clarify concepts and isolated problems; it does not complete an assignment, choose a

project for you, or pre-grade a submission.

Do not share quiz or midterm questions, answers, or screenshots with classmates or in public posts, including after an assessment closes. Direct any necessary assessment-content question to your mentor in a private Ed post; follow the [Quizzes and Exams rules](#).

For how to join an optional help session, see [Office Hours](#). For proctored-assessment incidents, follow [Honorlock](#) first, then send the private Ed report.

Honorlock

This page covers proctoring setup and technical incidents. See [Quizzes and Exams](#) for assessment content, permitted materials, and the midterm reference-sheet policy.

Prepare with Quiz 0

Before taking any graded quiz or the online midterm exam, open Quiz 0, answer its practice question(s), and complete the submission using the same computer, browser, camera, microphone, and internet connection you will use for that assessment. Confirm that Honorlock launches correctly and that your equipment and internet connection are working reliably. Quiz 0 is ungraded.

Use Quiz 0 to identify and resolve setup problems before starting a graded attempt. Do not start a graded quiz or the midterm exam with an unresolved problem. A successful practice attempt cannot guarantee that no later outage will occur.

You are responsible for your equipment, internet connection, and testing readiness. The teaching team is not responsible for problems with your personal setup. A technical problem does not automatically entitle you to a restart, extra time, a make-up, or credit for a quiz or midterm exam you could not complete. Unless an exception is officially approved, an uncompleted quiz or midterm exam receives no credit. For an incident during an attempt, follow [technical support instructions](#). For an accommodation or exceptional circumstance, use the [official approval and Hub-document process](#); contacting staff is not a substitute for that process.

Before each assessment, confirm that your camera and microphone work, disconnect external monitors and unauthorized devices, close communication, recording, remote-access, and AI applications, and launch from the Honorlock-enabled Canvas link.

Websites during an assessment

Do not browse websites during a quiz or the midterm. The only sites permitted in the proctored browser are Canvas/Honorlock and, **for the midterm only**, the authorized ML Class Hub link to your own reference sheet. Other Hub pages are not permitted. Do not use a second device or another browser to reach a blocked site. If the reference-sheet link does not open during the midterm, contact Honorlock live support immediately instead of bypassing the restriction, then report the incident to your assigned mentor in a private Ed post after the attempt.

Access-code or technical problem

For Honorlock problems, open your course in Canvas, select Honorlock, and use its built-in support/chat control. During an assessment, use the support control in the active Honorlock window and follow the proctor's instructions. Keep the support case number for any incident report. Do not leave the assessment to visit an external support website or use an unauthorized device.

If Canvas asks for an access code, do not request one from another student or the team. Exit before beginning, confirm Chrome and the Honorlock extension are active, then relaunch through the Honorlock-enabled Canvas link. Contact Honorlock live support if the problem continues.

If a problem occurs during an attempt, contact Honorlock live support immediately, follow its instructions, preserve the case number, and submit a private Ed post to your assigned mentor promptly afterward with the assessment name, time, and factual description. Do not open an unauthorized app or device to troubleshoot. A technical incident does not automatically grant another attempt; the team reviews the available records.

Assessment security and review

Do not bypass or interfere with Honorlock using a virtual machine, remote desktop, screen sharing, second device, unauthorized browser environment, or another method. Suspected violations are referred to the Georgia Tech Office of Student Integrity under Institute procedures. An automated Honorlock flag alone is not a misconduct finding; the team reviews relevant recordings and logs.

Approved disability accommodations are implemented with the Office of Disability Services. See [General Course Policies](#) for the approval and Hub-document process.

GRADING



Grading



Categories

Grading Categories

The required course components total 100%. There is no final exam. Teams can earn up to **5 additional course percentage points** for exceptional optional work beyond the required team project, for a maximum course score of 105%. The **bonus criteria** explain what qualifies.

Component	Course weight
Quizzes 1–14	15%
Midterm exam	10%
Individual project	35%
Team project	40%
Total	100%

Quiz 0 is required preparation but carries no course-grade weight. See [Quizzes and Exams](#) for assessment rules and [Project Breakdown](#) for the two project tracks. Project stages have these course-grade weights:

Individual project stage	Course weight
Midterm video and website update	10%
Final interview	15%
Final website	10%

Team project stage	Course weight
Proposal	7%
Midterm team interview (no video)	10%
Final team video (one per team; no interview)	10%
Final website and report	13%

The individual midterm website update is the work presented in the video, not an additional graded item. The individual final interview and completed website are separate graded components.

For the team final, the video assesses the team's explanation of its work; the website and report provide the full methods, results, and supporting evidence.

Canvas provides the grading entries. The teaching team enters one shared score for a team stage and records an individual score only when a member's participation warrants a different grade. The Hub remains the place for project work and team coordination.

Quizzes and Exams

Quizzes and Exams

This page covers assessment content, availability, and permitted materials. See [Grading Categories](#) for weights and [Honorlock](#) for proctoring setup and technical problems. There is no final exam.

Quiz 0 and graded quizzes

Before every graded quiz, complete the [Quiz 0 readiness check](#). Read the setup-responsibility and technical-incident rules there before starting.

Quiz 0 is the ungraded course-readiness check. Quizzes 1–14 are graded, completed individually in Canvas, and proctored with Honorlock. Each graded quiz is closed book and closed notes: no reference sheet, lecture tab, external website, generative AI, calculator, communication, or help from another person unless that quiz's Canvas instructions explicitly authorize a resource.

The regular quiz window is **Friday at 12:00 AM through Monday at 11:59 PM Eastern Time (ET)**. The published Canvas times control if an individual quiz has an announced exception. Canvas also shows duration, points, and attempts. Once an attempt starts, its timer runs continuously; leaving the browser does not pause it. Start with enough time to finish before closing.

Questions draw on assigned lecture videos, slides, written notes, annotations, explanations, and examples. Practice questions do not exhaust the graded bank. Do not share or discuss quiz questions or answers with other students or publicly, even after a quiz closes; direct a necessary question to your assigned mentor through a private Ed Discussion post.

Midterm exam

Online students: take the midterm online. Before starting, complete the [Quiz 0 readiness check](#) using your planned exam setup. The same equipment, internet, technical-incident, and approved-exception rules apply to the online midterm.

Campus students: take the midterm in class. Follow your instructor's in-class exam instructions and the announced date, time, and location.

The midterm is the only exam. It covers assigned material through dimensionality reduction (L14), inclusive. Study the assigned videos, slides, and written notes; practice quizzes are not an exhaustive list. The exam draws questions from a bank, so students can receive different questions. Do not assume another student's exam is identical to yours.

For online students, take the midterm individually in Canvas with Honorlock. Canvas publishes the online exam availability window, duration, and attempt limit. Once started, its timer runs continuously and the attempt must be completed in one sitting. Start with enough time to finish before the window closes.

Exam questions, answers, screenshots, and details must remain private. Do not discuss, copy, post, or share them with classmates, on Ed, or anywhere else, during or after the exam window. If you need to ask about a question or report a problem, contact your assigned mentor through a private Ed Discussion post; see [contact instructions](#). Do not send exam content to another student.

Midterm reference sheet

You may prepare one personal PDF of at most **two pages and 5 MB**. Typed, handwritten, and scanned pages are allowed. It must be legible at ordinary on-screen magnification; typed main content should be at least 9-point text. The reference sheet is not permitted for quizzes.

Upload and preview the PDF in [ML Class Hub](#), not Canvas. The Hub keeps one current file per student; you may replace it before starting your exam. Do not replace it during an active attempt. During the midterm, use only the authorized in-exam link to your Hub reference sheet. Other Hub pages, external websites, printed or locally saved copies, email, cloud drives, and another student's sheet are not permitted unless an approved accommodation says otherwise. If the authorized link does not open, contact Honorlock live support immediately, then report the incident to your assigned mentor through a private Ed post after the attempt; do not bypass the browser restrictions.

Accommodations and exceptional circumstances

For official accommodation and make-up procedures, use [General Course Policies — Accommodations and exceptional circumstances](#). A private Ed

message alone does not approve an exception. For an assessment technical incident, follow [Honorlock](#) immediately.



Projects: Individual and Team

Individual and Team Projects

These are **two separate projects**, each with its own website link in [ML Class Hub](#). Keep each link throughout the semester and update the appropriate site at its checkpoints. Dates are on the [course schedule](#); grade weights are in [Grading Categories](#).

Individual project

Work on your **own** using the dataset assigned in ML Class Hub. Follow the required analysis and website layout. Start with the [Individual Project guide](#).

Team project

Work with your **registered team** on a question and lawful dataset your team chooses. Use one team website from proposal through final. Start with the [Team Project guide](#).

Midterm and final: what to submit

There are **two live TA interviews**: the team midterm interview and the individual final interview.

Before attending either interview, read [Guidelines → Project Interview Rules](#).

Checkpoint	Individual project	Team project
Midterm	One unlisted YouTube video per student; no interview	Live team interview; no video

Checkpoint	Individual project	Team project
Final	Live individual interview; no video	One unlisted YouTube video for the entire team; no interview

Keep updating your existing project websites. The video and interview requirements do not replace the website deadlines.

Where to submit video links

Open [your projects in ML Class Hub](#) and select the correct course and checkpoint:

- **Individual midterm video:** each student saves their own unlisted YouTube link in their individual midterm project.
- **Team final video:** one team member saves one shared unlisted YouTube link in the team's final video project. Other members do not submit separate copies.

Use **Save or replace video link** on the relevant project page. Confirm that the saved link appears, test playback while signed out of YouTube, and keep the video available through grading. Choose **Unlisted**, not Private; anyone with the link can watch and reshare it. These are separate submissions: do not put a team video in the individual checkpoint.

The two websites, submissions, and review requirements are different. Use the guides above to find the instructions for the project you are working on.



Individual Project

Individual Project

Start with [Requirements](#) for the assigned work and checkpoint instructions.

- [Website Template](#): required structure and downloadable starter.
- [Grading Categories](#): the authoritative grade breakdown.
- [Course Schedule](#): checkpoint dates and deadlines.
- [Project Interview Rules](#): requirements for live interviews.

Requirements

Requirements

ML Class Hub will give you the **required dataset**. Do not choose a different one. The final prediction question or questions depend on that dataset and will be stated separately; this website template does not decide them. This is your own project, separate from the team project.

Start with the required website template

Create **your own GitHub Pages project website** and register its link in ML Class Hub. Keep that same link all semester. The word *website* below means **your project website**, not this course-instructions page.

Open the [rendered individual website template](#) to see the **required headings, subsection order, fill-in explanations, and example tables**. Use the download link near the top of that page to save the complete Markdown starter (`.md`) for your own website. Keep its numbered headings and fill them with your own work. The assigned dataset and target groups come from ML Class Hub; sample text is not a result to copy.

Midterm: find patterns without the answers

First, prepare the assigned data. Then use at least one way to keep fewer original features or create fewer new features. Run a **clustering method** twice: once using all features and once using the smaller feature version. Show the groups, use suitable class-taught measures that do not need labels, and explain what worked or failed. The [template's midterm section](#) walks through every subsection.

Submit **one 10-minute unlisted YouTube video link** in the individual project area of ML Class Hub. Explain your work and a result in the video. There is no individual TA interview at midterm; the team midterm interview is a separate requirement and needs no video. Open [your projects in ML Class Hub](#), select

your individual midterm checkpoint, and save your unlisted YouTube link. Test playback while signed out and keep it available through grading.

Final: predict the answers the course assigned

Continue on the **same website**. Once the course states what to predict for the dataset, write those questions on your project website in plain language. Do not assume that ML Class Hub has already supplied answer groups or a ready-made experiment.

- **CS 4641 · Undergraduate** : run **two** feature-selection/reduction methods and **two** prediction models.
- **CS 7641 · Graduate** : run **three** feature-selection/reduction methods and **three** prediction models.

For **each** prediction question the course states for the dataset, test each prediction model with all features and with each smaller feature version. Use the same training, validation, and test examples across models so the comparison is fair. The [template's final section](#) gives the exact layout and an example results table. Explain mistakes and limits, not only your best score.

Final TA interview and deadline

The individual final is a live interview. **No individual final video is required.**

Show your current website and code in the final TA meeting. Be ready to explain your data, method choices, comparisons, mistakes, and what you personally did. These are common question areas, **not a complete question list**; the TA may ask other relevant questions or request a small demonstration. There is no extra early website submission. You may keep improving the website until the final deadline on the [course schedule](#).

Before attending, read the [Project Interview Rules](#), which apply to both the team midterm and individual final interviews.

Individual Project Website Template

Individual Project

[Download the complete Markdown starter \(.md\)](#)

Open the downloaded file in a text editor and replace the instructions with your own work. If your browser displays the file instead of downloading it, save the page as a `.md` file.

This is the required layout for **your own individual project website** (the GitHub Pages link you register in ML Class Hub). This course page is only the example; your website will contain **your** answers, figures, and results. Keep the numbered headings and subsections below in this order. Replace the instructions under them with your own writing. At midterm, complete Sections 1–3 and mark final-only work as planned. By the final deadline, complete every section. Do not copy an example result as if it were yours.

ML Class Hub will give you the dataset. It does not supply a ready-made analysis or results. Use the dataset the course gives you; do not substitute another one. The final prediction questions depend on that dataset and will be stated separately. This template does not decide them.

1. Project question

1.1 What am I trying to learn?

State the question in plain language. Say why someone would care about the answer. For example, if the assigned data are images, a question might be whether useful groups can be found without labels and how well the images can later be classified. This is only an illustration, not an assigned question or dataset.

1.2 What have I completed so far?

At each checkpoint, briefly say what you have actually finished and what is still planned. **Do not describe a planned experiment as a result.** Update this short status when you update the same website.

2. The data provided by the course

2.1 What data did I receive?

Give the dataset name and the course-provided download or information link. Say how many examples it contains, what one example looks like, and what information (features) is available for each example. Explain missing values or data problems you noticed and what you did about them. If the course gives restrictions on using or sharing the data, follow those instructions; do not publish the data itself unless allowed.

2.2 What will I try to predict at the final?

Write the prediction question or questions the course specifies **for the dataset you receive**. State the input and the answer your model will predict. The course will post the dataset-specific prediction task in ML Class Hub. If it has not been posted yet, leave this subsection for the final update; do not invent a task or fill it with a placeholder.

2.3 How did I separate the data?

State how many examples you put in the training, validation, and test sets. Do not let the model learn from test examples or use their answers while choosing a model; that is data leakage. When comparing models, use the same saved test examples for each one.

3. Midterm: finding patterns without using the answers

Complete this section and the separate midterm video required in ML Class Hub.

3.1 How did I prepare the data?

Explain what you changed before running the method, such as handling missing values or scaling numeric features. First use all available features, then compare with a reduced-feature version. If a preparation step learns from data, fit it on training examples only.

3.2 How did I choose or reduce features?

Use at least **one** method that reduces the amount of information given to clustering. It may keep fewer of the original features or make a smaller set of new combined features. Name the method you used, say how many features remain, and explain your choice in plain language. Do not use known answers (class labels) to choose or train this midterm method.

3.3 What groups did the clustering method find?

Name the clustering method and its important settings. Run it twice: once with all features and once with the reduced-feature version from Section 3.2. State how many groups it found or was asked to find. Show examples or a figure that helps a reader see what the groups contain.

3.4 How did I judge the groups?

Use one or more ways **taught in class** to judge the groups **without looking at known class answers**. Name what you chose, give the actual result for grouping with all features and with fewer features, and explain whether that result is better or worse and why. No particular score is required in this template. Show a figure or a few examples from the groups and explain what they reveal. Do not show only a number without interpretation.

If the dataset has reliable known answers, you **may** compare them with the finished groups afterward as an extra check. This is optional. Do not invent answers, and do not use them to make or tune the groups. Also describe at least one group that makes sense and one result that is confusing or unsuccessful.

3.5 What did I learn, and what will I try next?

Summarize the main finding, one limitation, and the next experiment. Explain weak or unexpected results honestly. Your midterm video link is submitted in ML Class Hub, not as a new website.

4. Final: predicting the assigned answers

Keep Section 3 on the website and add the supervised-learning work below.

4.1 Which methods and prediction models did I run?

- **CS 4641 · Undergraduate** : run **two** different feature-selection/reduction methods and **two** different prediction models.

- **CS 7641 · Graduate** : run **three** different feature-selection/reduction methods and **three** different prediction models.

List the methods and models you **actually ran**. A feature method can carry over from Section 3 if you use it again here. Explain what each does in your own words and why you chose it. A name alone is not enough.

4.2 Which comparisons did I make?

For every prediction question the course specifies for the dataset, run each required prediction model first with **all features** and then with **each** of your required smaller feature versions. Change only the feature version or model being tested; keep the examples used for training, choosing settings, and final testing the same. Prepare the data using training examples, choose settings using validation examples, and use test examples for the final check. This lets a reader tell whether a different model or fewer features really helped.

4.3 What were the results?

For each prediction question, show the results for **every model you ran** with all features and with each smaller feature version. Use the **real names** of your methods and your **actual scores** in a clearly labeled table or chart. Explain what the score measures, whether higher or lower is better, and what mistakes the model made. Do not show only your best result. There is no placeholder table to copy: its rows and names must come from the dataset and methods you actually used.

4.4 What worked, what failed, and why?

Compare results that answer **the same prediction question**. Explain where using fewer features helped or hurt, which answers the model confused, and whether one method took much longer to run. Show at least one concrete mistake or failure. A small score difference alone does not prove that one method is always better. State what you cannot conclude from these experiments.

5. Conclusion

Answer the question from Section 1 using your actual results. State the main limitation and one useful next step. Do not claim more than the evidence shows.

6. References

List the sources you actually used and cite them where they support a claim, method, or dataset description. Do not invent references.

7. AI-use disclosure

If you used generative AI for this project, name the tool, what you used it for, how much of the work it affected, and what you checked yourself. Follow the [course AI policy](#). If you did not use it, say so.



Team Project

Team Project

Choose the guide for your current stage:

1. [Proposal](#)
2. [Midterm Review](#)
3. [Final Report and Video](#)

Resources

- [Website Template](#): report structure and downloadable starter.
- [Topic and Project Example](#): a worked example across stages.
- [Grading Categories](#): the authoritative grade breakdown.
- [Course Schedule](#): checkpoint dates and deadlines.
- [Project Interview Rules](#): requirements for live interviews.
- [Optional Bonus](#)
- [Awards and Examples](#)
- [General Course Policies](#)

Topic and Project Example

Choosing a Team Project Topic

Create something that matters beyond this course

A central goal of the team project is to develop your creativity through collaboration. Bring together your different backgrounds, interests, and experiences. Share ideas, brainstorm openly, question assumptions, and make room for every member to contribute. Connecting perspectives from different fields can reveal a useful question or approach that none of you would have found alone.

This course also gives you an opportunity that can be hard to find later: a group of people coming together because they are curious, want to learn from one another, and care about creating something worthwhile. Your motivation can be the shared purpose of the work itself, without needing an immediate financial payoff. Make the most of this time to explore an idea you believe in, encourage one another, and build something you are proud to share.

Think beyond completing an assignment: who could benefit from your work, what problem could it help solve, and how would you measure whether it helps? Your project could become the starting point for a research paper, a startup, an open-source tool, or a practical solution that improves other people's lives. Start with a focused, achievable version of that idea and use experiments and evidence to decide what to pursue.

Be willing to explore unexpected ideas, learn from unsuccessful experiments, and build on one another's strengths. The aim is to combine imagination with sound machine learning and honest evaluation. **These possibilities are aspirations, not additional grading requirements:** you do not need to publish a paper, launch a company, or achieve commercial success to earn full credit. Follow the [project requirements](#) and [grading criteria](#).

One example across stages

Suppose a team has customer reviews and asks “**Can we predict whether a review is positive or negative, and what themes appear in the reviews?**” This is only an example; teams may choose other questions, datasets, and algorithms.

1. **Proposal:** describe the data and plan to try a majority-label baseline, logistic regression, and a decision tree to predict positive/negative labels. A CS 7641 team also plans a method such as K-means to group reviews without using labels. Explain what each method would help answer; no results are required yet.
2. **Midterm:** a CS 4641 team has run the baseline and at least one classifier, and shows results and a limitation. A CS 7641 team has also run a grouping method and shows what the groups reveal. Show real work, even if results are weak.
3. **Final:** compare the baseline, logistic regression, and decision tree on the **same prediction task** using an appropriate test measure. The CS 7641 team also reports the grouping results separately. **Do not compare clustering and classification by pretending they have the same accuracy measure;** they answer different questions. Explain the tradeoffs, errors, and what was learned.

The named algorithms and review dataset are illustrative, not required. A different valid baseline, model family, or dataset can satisfy the requirements when the team explains its choices and results.

Choose an ambitious but feasible topic

A strong topic has a clear question, data you can lawfully use, a baseline, an evaluation plan, and time for error analysis. Possible domains include language, vision, health or scientific data, recommendations, time series, security, robotics, and responsible AI. Simply calling a model API, using a fashionable dataset, or promising high accuracy is not enough. Ask what your project will teach someone that a straightforward baseline cannot.

Originality can come from a new question, data source, application setting, careful comparison, method adaptation, evaluation under realistic constraints, or a usable system. Novelty is encouraged, **not required for full credit**. A careful, reproducible study can be excellent even when its models are established. The [past awardees](#) offer inspiration, not templates to copy; some older Georgia Tech links require sign-in.

Treat the project as portfolio-quality work: document your own contribution, methods, evidence, limitations, and code so you can explain it on a résumé or in an interview.

Team Proposal

Team Project Proposal

Write the proposal on the **same team website** you will use through midterm and final. Register the team and link in [ML Class Hub](#) by the [course schedule](#) deadline. This is for the team project, not the individual project. No proposal video or separate proposal website is required.

Choose a feasible question

Identify a meaningful ML question, a lawful and obtainable dataset, a simple baseline, and a measurable outcome. The [past awardees](#) can inspire the scope, but do not copy their questions or work. Novelty can come from the question, data, application, method adaptation, realistic constraint, or rigorous comparison; it is **not required for full credit**. Investigate relevant prior work before claiming a contribution, and plan a smaller fallback if data or a complex method fails.

Expectations by course level

- **CS 4641 · Undergraduate** : name at least three ML models or methods the team plans to try. Include at least one supervised model.
- **CS 7641 · Graduate** : name at least three ML models or methods the team plans to try. Include at least one supervised model and one unsupervised method.

At this stage, name the planned methods, the question each one will answer, and what evidence you will use to judge the results. You do not need to have run them yet. See the [worked example](#) and [plain-language definitions](#).

Proposal sections

Use these headings so the same report can grow at midterm and final:

1. **Introduction and related work:** explain the question, motivation, relevant work you have read, and what your team plans to investigate.
2. **Problem definition and data:** state the specific question, the dataset you plan to use, how you will obtain it, and any access, license, or privacy limits.
3. **Methods and baseline:** name the simple comparison and at least three ML methods you plan to try; explain what each would tell you. Results are not required yet.
4. **Evaluation and results:** say how you plan to divide the data, measure success, check errors, and compare methods answering the same question. Mark all expected results as planned, not completed.
5. **Discussion and limitations:** briefly identify risks, likely data or method limitations, and a fallback if your first plan fails.
6. **Conclusion and next steps:** state what a useful answer would look like and the next experiment you will run.
7. **Project plan and member contributions:** include milestones, owners, and a Gantt chart for the rest of the semester. The [sample Gantt chart](#) is optional; a clear equivalent is fine. Name each member's contribution to this proposal.
8. **References and AI-use disclosure:** cite sources you actually read and disclose planned or actual generative-AI use under the [course AI policy](#).

Keep the proposal narrative **under 800 words**; references, the Gantt chart, and the contribution table do not count. A short, specific plan is more useful than a long speculative report. **Keep the same main project topic and core research question from the proposal through the midterm and final.** You may refine or replace your methods, algorithms, models, and experimental approach as you learn more, review results, or receive TA feedback. These changes must continue to address the original project question rather than introduce a different subject or project. In each website update, briefly explain what changed, why it changed, and how it supports the original objective. The proposal is a starting plan, not a commitment to use every initially proposed method; the course-level method requirements still apply.

Team Midterm Review

Team Midterm Website Review and TA Meeting

Update the **same team website** before the meeting with your assigned TA. This is the team midterm checkpoint, not a separate report or video. The [course schedule](#) gives the review window; the individual project's video is a separate requirement.

Update the proposal, do not start over

Keep the proposal's section headings and turn plans into evidence:

1. **Introduction and related work:** refine the question and explain any justified change since the proposal.
2. **Problem definition and data:** show the data obtained, preparation completed, access or privacy constraints, and data limitations.
3. **Methods and baseline:** distinguish methods you have run from methods still planned. Show a working simple comparison or early experiment where feasible and explain your choices.
4. **Evaluation and results:** show early quantitative measures and useful figures, explain what they mean, and identify unfinished experiments. Do not claim results you have not obtained.
5. **Discussion and limitations:** explain an early success, failure, or obstacle; identify what the evidence cannot yet establish.
6. **Conclusion and next steps:** summarize what you have learned so far and the next experiments before final.
7. **Project plan and member contributions:** update the [proposal Gantt chart](#) and identify each member's work, blockers, and next responsibilities.
8. **References and AI-use disclosure:** update sources and disclosed AI assistance actually used.

- **CS 4641 · Undergraduate** : run at least one ML model on your data, show its result with a suitable measure or figure, and compare it with a simple baseline.
- **CS 7641 · Graduate** : run at least one supervised model and one unsupervised method on your data. Show a result for each with suitable measures or figures, and explain what you learned.

For both course levels, show the actual output, say what it means, explain what did not work, and identify what you will improve. A model listed in a plan but not run does not meet this checkpoint. See the [worked example](#).

During the meeting, show the current website and working evidence, discuss blockers, and answer questions. Common questions may cover the research question, data, baseline or first experiment, evaluation plan, current result or obstacle, and each member's contribution. These are examples, not a complete question list. The TA may ask other relevant questions about any part of the team's work shown live. Every member should attend and understand the whole project. The TA records feedback and next steps. Apply it before the [final report](#). Methods not yet taught are not required at this checkpoint.

Attendance and individual participation

The midterm review is worth 10 course percentage points: 7 for the team website and progress demonstrated at the meeting, and 3 for each member's own discussion of the work. The TA gives the team one shared website/progress score and records an individual discussion score when participation differs. The questions listed above are examples; the TA can ask other relevant questions about the project.

If an Institute-approved accommodation prevents attendance, contact your assigned mentor through a private Ed Discussion post (see [contact instructions](#)). Staff will arrange an appropriate alternative; teammates do not need to approve an accommodation. Do not post accommodation documents or details in team chat.

For another scheduling conflict, tell the team and TA before the meeting. If every other team member agrees in the Hub team chat, the absent member can post a link to a 10-minute unlisted YouTube explanation of their contribution in that chat before the meeting and respond to the TA's follow-up questions. The shared team score is unaffected. Without attendance or one of these alternatives, the member does not earn the 3 individual discussion points; demonstrated team work still receives its shared score.

Before attending, read the [Project Interview Rules](#), which apply to both the team midterm and individual final interviews.

Team Final Report

Team Final Report, Website, and Video

Complete a **paper-style report on the same team website**, with supporting code and reproduction instructions. Submit **one unlisted YouTube video for the entire team**. There is **no team final interview**. See the [course schedule](#) for the separate deadlines for the team final video and completed website/report. Keep the same ML Class Hub team link. The individual final is separate; see [Individual Project](#).

Use the [rendered team website template](#) as the minimum report structure, including the required interactive visualization. You may add useful sections and features beyond it while keeping the core headings.

Complete the same report

Develop the proposal and midterm sections into a coherent final paper:

1. **Introduction and related work:** state the question, motivation, verified prior work, and specific contribution or comparison. Do not claim novelty without evidence.
2. **Problem definition and data:** describe the data source, license, preparation, data division, responsible use, and limitations.
3. **Methods and baseline:** explain the methods you actually ran, the simple comparison, and enough detail for a fair comparison.
4. **Evaluation and results:** present measures, tables or figures, meaningful comparisons, and the required interactive visualization of your own data or results.
5. **Discussion and limitations:** explain successes, failures, tradeoffs, alternative explanations, and what the evidence does and does not establish.

6. **Conclusion and next steps:** answer the original question as far as the evidence allows and identify the next useful experiment.
 7. **Project plan and member contributions:** show the final Gantt-chart record or milestone summary and each member's actual contribution.
 8. **References and AI-use disclosure:** cite reviewed work and disclose tools, purpose, and scope of AI assistance. Be ready to show the project code to the TA if asked. Do not publish restricted data.
- **CS 4641 · Undergraduate** : run and compare at least three different ML models or baselines, including at least one supervised model.
 - **CS 7641 · Graduate** : run and compare at least three different ML models or methods. Include both a supervised model and an unsupervised method.

A simple baseline can count as one of the three only if the team actually runs it, reports its results, and compares models answering the same question with it. Show the results of all required methods and explain what the differences mean. Supervised prediction and unsupervised grouping answer different questions, so report each with suitable evidence rather than forcing one accuracy comparison across both. No external clustering measure is required when reliable labels are unavailable. See the [worked example](#).

The final team website must include an interactive visualization that lets a reader explore the team's own data or model results; a static plot alone does not satisfy this requirement. A working application beyond that visualization may be considered for the [optional bonus](#).

The report is not subject to the proposal's 800-word limit. A working demo can help when relevant, but a polished interface does not replace valid ML evidence.

Team final video: one link for the whole team

Explain the project question, data, baseline, methods, meaningful results, limitations, and each member's actual contribution. Show the website, relevant code, and evidence that supports your conclusions. Every member should understand the whole project.

1. Upload the team's video to YouTube and choose **Unlisted**.
2. Open [your projects in ML Class Hub](#) and select the **team final video** checkpoint for your course.
3. One team member saves the video link on behalf of the whole team. Do not submit a separate video for each member.

4. Confirm that the saved link appears and plays while signed out. Keep it available through grading. Anyone with an unlisted link can watch and reshare it.

The video and final website/report are separate graded components. See [Grading Categories](#) for their weights. You may continue improving the website until the final website deadline in the [course schedule](#). The individual final interview remains required, and the team midterm remains a live interview with no video.

Team Project Website Template

Team Project

[Download the complete Markdown starter \(.md\)](#)

Open the downloaded file in a text editor and replace the instructions with your own work. If your browser displays the file instead of downloading it, save the page as a `.md` file.

This is a starting layout for **your team's own project website**, not a topic or result to copy. Keep the main headings so your TA can find the required work. Your team **may add subsections, better explanations, interactive features, and a different visual design**. Use this same website for the proposal, midterm review, and final. At proposal, write what you plan; at midterm, show what you have done; at final, revise it to show completed work. The team chooses a feasible question and lawful dataset, unlike the individual assignment where the course supplies the data.

1. Introduction and related work

1.1 What question are we trying to answer?

State one specific question, why it matters, and what would count as a useful answer. Say when the question changes and why.

1.2 What have others already done?

Describe a few relevant sources you actually read and what your project will test, compare, or build beyond them. You do not need to claim a new research discovery.

2. Problem definition and data

2.1 Which data did we choose, and can we use them?

Give the dataset name, a link to where the team obtained it when possible, the number of examples, the information each example contains, and whether answers (labels) are known. State any license, privacy, or sharing restriction that applies. If the first dataset was unavailable, explain the change.

2.2 How did we prepare and divide the data?

Explain what the team cleaned or changed. State which examples were used to train methods, which helped choose settings, and which were saved for final testing. Explain how you kept test information out of training.

3. Methods and baseline

3.1 What simple result are we trying to beat?

Name the simple baseline you will compare against and show its actual result by midterm or final. Compare it with methods addressing the same question.

3.2 What other methods did we try?

For each method, explain why you chose it and whether it is **planned**, **in progress**, or **completed**. At proposal, name at least three methods or models you plan to test; results are not required yet. By final, run and report at least three different methods or models in total. A baseline counts only when actually run and compared. CS 4641 must include a supervised model; CS 7641 must include supervised and unsupervised methods. See the [course project requirements](#) for details.

4. Evaluation and results

4.1 What did each method produce?

Show real numbers, tables, or figures and explain what they mean. At proposal, write what you plan to measure. At midterm, show your first results, even when weak. At final, show the completed results. Methods answering different questions should not be forced into one comparison: for example, a prediction model and a grouping method need different explanations and may need different measures.

4.2 Required interactive visualization

Include **at least one useful interactive visualization** of the team's own data or model results by the final deadline. A reader should be able to change something—such as a filter, class, feature, threshold, or model setting—and see a relevant chart or result change. JavaScript is one way to build this. Explain what the interaction helps the reader understand. A static image or decorative animation alone does not meet this requirement.

5. Discussion and limitations

Compare results that answer the same question. Explain at least one success, one failure or surprising case, and an important tradeoff. Describe data gaps, possible bias, uncertainty, privacy concerns, and other limits that affect how someone should interpret or use the result. Say what the team observed separately from what it guesses.

6. Conclusion and next steps

Answer the question from Section 1 as far as the evidence allows. Name one next experiment that would be worth doing.

7. Project plan and member contributions

Include a Gantt chart or simple timeline showing milestones and who owns them. At each checkpoint, update it to show what actually happened. Name each member's work and next responsibility; all members must understand the whole project.

8. References and AI-use disclosure

Cite sources the team actually used. If generative AI helped, name the tool, purpose, scope, and what the team verified under the [course AI policy](#). If none was used, say so.

Optional bonus work (not required)

The required interactive visualization is **not bonus work**. If your team requests bonus credit, list the specific bonus items completed, show the evidence, and identify who did each item. See the **Optional Team-Project Bonus** page on the course website for the five checkable items and their limits. Bonus credit is not automatic.

Bonus

Optional Team-Project Bonus

The required course components total 100%. A team can earn up to **5 additional course percentage points**, making 105% the maximum possible score. The individual project has no separate bonus. The same five items apply to CS 4641 and CS 7641 teams, but each item must be **additional to the required work for that course level**. A team can earn full required-project credit without any bonus work.

How to request bonus points

Add an **Optional bonus work** section to the final team website. For each item claimed below, include the requested evidence and name the members who did it. Show the work during the final TA interview if asked. There is no separate upload or deadline. Each completed item earns **+1 course percentage point**; incomplete or unverified items earn **0**. The same data, experiment, visualization, or software feature cannot be counted for two items. At most five points can be earned.

Five checkable bonus items

1. **Collect new real-world data (+1)**: gather observations yourself for this project rather than downloading an existing dataset or creating synthetic, augmented, or AI-generated examples. On the website, state what was observed, how and when it was collected, how many usable observations resulted, and which members collected them. Show a small permitted sample or a collection log. Follow applicable consent, privacy, and data-use rules; do not publish restricted information. Cleaning or augmenting an existing dataset alone does not count.
2. **Test design choices (+1)**: run at least two controlled comparisons beyond the required model comparison, changing one choice at a time while keeping the same data split and evaluation measure. For example, test two feature sets and two training-set sizes. Show a table with the choices and

results, and explain what each comparison taught you. Merely running more models or listing settings without results does not count.

3. **Test on a second independent dataset (+1):** evaluate a completed method on a second, separately obtained dataset that addresses the same question; a re-split, augmentation, or synthetic copy of the first dataset does not count. State its source and differences from the first dataset, show the result on both datasets using a suitable measure, and explain whether the conclusion changed. The dataset used to claim item 1 cannot also be used to claim this item.
4. **Build an enhanced interactive explanation (+1):** go beyond the required single interactive visualization. Provide at least three meaningful controls (for example, choose a model, choose a feature, and filter a data group) that update actual project data or model results. Show a comparison between at least two methods or groups, explain what readers learn from it, and demonstrate that the controls work. Extra colors, a static plot, or decorative animation do not count.
5. **Build a working model-based tool (+1):** provide a separate usable workflow beyond the report and its visualization. A user must be able to supply a new input, run the team's trained model or method, and see an explained output. Show one example input and output and demonstrate the tool during the final interview. A mockup, static page, or visualization of existing results alone does not count.

Privacy, honest reporting, citations, and disclosure of generative-AI use are required for **every** project; they do not earn bonus points by themselves. **Project awards** are separate recognition.

Awards and Examples

Project Awards and Examples

These awards recognize exceptional **individual and team projects**. Awards are distinct from course grades and the [optional team-project bonus](#); an award is not needed for full credit or bonus credit. Every completed project is eligible for consideration; no separate award application is required. The final website and interview are the evidence reviewed. See [Grading Categories](#) for the course-grade weights.

What makes a project stand out

- A clear, important question and an honest account of related work.
- Correct methods, a meaningful baseline, valid evaluation, and careful failure analysis.
- A distinctive insight, application, dataset, method adaptation, rigorous comparison, or useful working system. Originality is not just a new model name or a polished interface.
- Reproducible code and evidence, responsible data use, verified citations, and clear disclosure of contributions and generative-AI assistance.
- A website and explanation that let another person understand what was attempted, what the evidence supports, and what remains unresolved.

The instructional team reviews completed websites and TA interview evidence after the final website deadline. It can recognize different strengths, such as research potential, applied impact, reproducibility, responsible AI, or effective communication; no category or number of awards is guaranteed. Website presentation alone does not overcome invalid experiments or unsupported claims. Public showcasing of a student's work requires permission and a privacy/licensing review; merely submitting a project does not grant permission to publish it elsewhere.

Explore earlier projects

The following links were already listed in the course's historical award archive. They are **examples for inspiration, not approved templates or verified current links**. Many Georgia Tech GitHub Pages links require Institute sign-in, and access can change. Do not copy another team's text, code, data, or experimental design. The placeholder Spring 2026 entries have been removed.

▼ Project Awardees for Fall 2025

Out-of-Stock Recommendation System

- **Project Website:** <https://github.gatech.edu/pages/Group44-OOS-Recommendation/OOS-Product-Recommendation-Engine/>
- **Contributors:** Arihanth Jayavijayan, Riya Bharathwaj, Nirmal Francis Xavier, Malhar Jadhav

Predicting Life Expectancy with Interpretative Machine Learning

- **Project Website:** <https://github.gatech.edu/pages/frahman39/CS7641-Group38/>
- **Contributors:** Amit Badoni, Allanda (Lonnie) Kriener, Tahsin nabi, Avinash Palliyil, Farhan Rahman

invisiGAN

- **Project Website:** <https://github.gatech.edu/pages/jmaayah3/invisiGAN/>
- **Contributors:** Jameel Maayah, Caleb Rieck, Srikar Satluri, Ajinkya Argonda, Anish Vallabhaneni

Wheel-Leg Robot Robust Locomotion through Deep Reinforcement Learning and Clustering

- **Project Website:** <https://github.gatech.edu/pages/spolisetti6/ml-group-81/>
- **Contributors:** Zimeng Chai, Collin Agarwal, Sasanka Poliseti, Trent Doiron, Nathan Donagi

▼ Project Awardees for Summer 2025

Inverse Kinematics Approximation Using Machine Learning

- **Project Website:** <https://github.gatech.edu/pages/jmerrick6/InverseKinematics/FinalReport.html>
- **Contributors:** Terry Barrigah, Joao Fonseca, Jackson Thomas Merrick, Anand Nagpurkar, Yash Tahiliani

▼ Project Awardees for Spring 2025

CEGS-LIME: Clustering-Enhanced, Graph-Search LIME for CNN Explainability

- **Project Website:** <https://github.gatech.edu/pages/vkulkarni46/CS-7641-ML-Project/>
- **Contributors:** Vineet Kulkarni, Amogh Palasamudram, Ayush Panda, Khooshrin Pithawalla, Krishna Ravishankar

Advanced Quadrotor Control & Stabilization

- **Project Website:** <https://github.gatech.edu/pages/xyang626/ML-7641-Quadrotor-Group/docs/index.html>
- **Contributors:** Dyllon Preston, Evan Dodani, Dennis Anthony, Ajay Tak, Xinyi Yang

Restaurant Review Prediction

- **Project Website:** <https://github.gatech.edu/pages/agoli6/restaurant-review-prediction/>
- **Contributors:** Venkata Goli, Jack Hayley, Vasilisa Lyubimova, Prajval Manivannan, Roderic Parson

Learning UAV State Estimator

- **Project Website:** <https://github.gatech.edu/pages/adeb40/ML-Project-Group-60/>
- **Contributors:** Aditya Deb, Benjamin Monnig, Hunter Kuperman, Tristan Thakur, Xiaomeng Ye

RESOURCES



Resources

Collection

GT Resources

For any resources related to Student Engagement and Wellbeing such as Dean of Student's contacts, Library, etc., please refer to this [LINK](#)

Class Resources

Main Course Reference

The main reference for this course is *Pattern Recognition and Machine Learning (PRML)* by Christopher M. Bishop. The course is based primarily on this book, with lecture materials and assigned readings providing additional coverage. A free [official PDF](#) is available, so purchasing a printed copy is not required. Complete the readings assigned for each class.

Books and Additional References

- [Pattern Recognition and Machine Learning \(PRML\)](#), by Christopher M. Bishop — **main course reference**; [free official PDF](#).
- [Learning from Data](#), by Yaser S. Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin — official book website.
- [Machine Learning](#), by Tom Mitchell — author's book website and resources.
- [Data Mining: Concepts and Techniques](#), by Jiawei Han, Micheline Kamber, and Jian Pei — third-edition companion website.
- [The Elements of Statistical Learning](#), by Trevor Hastie, Robert Tibshirani, and Jerome Friedman — authors' website with a free PDF.
- [Deep Learning](#), by Ian Goodfellow, Yoshua Bengio, and Aaron Courville — free online book.

The additional books are supplementary references, not required purchases.

Dataset Ideas

May need API, or scraping – thanks to [Polo](#) and everyone who contributed with suggestions to these datasets:

- [HuggingFace Datasets](#). [Thanks to Xuhui Zhou] Popular dataset-hosting website for machine learning, especially for natural language processing problems. The unified API is convenient for training models.
- [Google Dataset Search](#)
- [Data Commons](#)
- [Kaggle public datasets](#)
- [Awesome Public Datasets](#)
- [NYC TLC trip records](#): taxi and for-hire vehicle trip data.
- [Georgia Tech's campus data \(has APIs\)](#): campus places and building information.
- [Data.gov](#): U.S. Government's open data
- [IPEDS data](#): Postsecondary education data from National Centre for Education Statistics
- [Bureau of Labor Statistics data](#)
- [Wikidata](#): structured knowledge with query and download options.
- [Yelp Open Dataset](#): reviews and business data for educational use; check the dataset terms.
- [OpenAlex](#): scholarly works, authors, institutions, and citation data.
- [Numerous APIs from Google](#) (e.g., YouTube)
- [Zillow Research housing data](#).
- [Graph datasets: SNAP](#).
- [Movies data: IMDb non-commercial datasets](#).
- [List of lists of datasets for recommendations](#).
- [Million Song Dataset](#): audio features and metadata; see the repository for available files.
- [Dataset about soccer games, players, clubs](#). Check the site's access and reuse terms before collecting data. For a soccer player: transfer history, performance, nationality, birth date, etc. For a soccer club: performance, squad, etc.
- [Nasdaq Data Link](#): financial and economic data; availability and access terms vary.
- [UCI Machine Learning Repository](#): datasets for classification, regression, and other tasks.
- [Amazon AWS Public Data Sets](#)
- [KDD Cup](#): data-mining competitions.
- [Bibliographic data: DBLP XML downloads](#).

- [Retrosheet](#): MLB statistics (Game/Play logs)
- [Classification datasets](#)
- [NOAA NCEI data products](#): environmental and geophysical data.
- [Pew Research Center datasets](#).
- [Academic torrents](#) (terabytes)
- [New York Times Article Search API](#).
- [Case Western Reserve Geotechnical Database](#).
- [Data Science Initiative – Microsoft Research](#) has various datasets and access to tools that can aid in data science research

Other resources, such as machine learning toolboxes and datasets, will be provided throughout the course.