PRINTABLE COURSE PACKET

Machine Learning CS 4641 & 7641

Summer 2026

PACKET SUMMARY

12 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 Overview • Instructors and TAs (Max) • Course Schedule (Mahdi)

GUIDELINES

2 pages selected

General • Office Hours

GRADING

6 pages selected

Categories • Project Breakdown • Proposal • Midterm Checkpoint • +2 more

RESOURCES

1 page selected

Collection

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 learningalgorithms
- Help you learn to apply tools for real dataanalysis problems
- Encourage you to do research in data science andmachine 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 on team projects in a remote environment
- Conducting peer evaluation in a constructive format
- Communicating technical content in a concise and effective manner

COURSE INFO

Instructors and TAs

Included: Max

Instructors and TAs

Instructors

Max Mahdi Roozbahani
mahdir@gatech.edu
https://mahdi-roozbahani.github.io/

COURSE INFO

Course Schedule (Mahdi)

Schedule

Important
All deadline and due dates in this course will be at 23:59 EST.

Scroll horizontally to see the full schedule table on mobile devices

Week	Dates	Topics	Homework	Quizzes	Project	Readings
1	5/18 - 5/22	• *Course Overview (L1) • Linear Algebra (Notes, L2)		Q0 - L1 | Q1 - L2		• GT Honor Code; • Heilmeier catechism; • Visual Information Theory by Chris Olah; • GitHub Pages; • YAML Configuration; • Overleaf for GT students; • Linear Algebra Review by Zico Kolter;
2	5/25 - 5/29	• Memorial Day Holiday • Prob and Stats (Notes, L3)	• A1 out 5/25	Q2 - L3 | Syllabus Quiz	• Project team composition due 5/29	• Independence, Coorelation, Orthogonality • Correlation vs Covariance; • The Differences Between Data, Information and Knowledge • Cross Entropy as loss function • More about Cross Entropy and KLD; • Probability Theory Review by Andrew Moore;
3	6/1 - 6/5	• Info Theory (Notes, L4) • Optimization (L5) • *Toolbox - P1/P2 (L6)		Q4 - L4, 5		• Debugging Common Errors in NumPy; • NumPy Tutorial; • Matplotlib Tutorial; • seaborn: statistical data visualization; • KKT for inequality constrained optimization; • Why Cross Entropy over MSE for Classification; • Gradient Descent short video; • Matplotlib Tutorial; • NumPy Tutorial;

Week	Dates	Topics	Homework	Quizzes	Project	Readings
4	6/8 - 6/12	• Clustering & K-Means (Notes L7) • Hierarchical Clustering (Notes L7) • GMM - Part 1 (Notes L8)	• **A1 due 6/12** • A2 out 6/12	Q5 - L7 | Q6 - L8		• Curse of dimensionality (Euclidean space example); • Jupyter Notbook (Kmeans and DBSCAN); • Understanding the concept of Hierarchical clustering Technique; • Dendrogram Visualization; • GitHub Student Application
5	6/15 - 6/19	• GMM - Part 2 (Notes L8) • DBSCAN (Notes L9) • Clustering Eval (Notes L10)		Q7 - L9, 10	• Project proposal due 6/19 • Peer Evaluation 6/24	• Jupyter Notbook (Kmeans and DBSCAN); • KDE interactive visualization; • KDE sampling; • KDE SKLearn and sampling; • Jupyter Notebook Kernel Density Example;
6	6/22 - 6/26	• Dimension Reduction (Notes L11) • Linear Regression (Notes L12)		Q8 - L11		• Image reconstruction using PCA; • Feature extraction using PCA; • PCA for images; • PCA as linear combination of features; • PCA and Linear Discriminant Analysis;
7	6/29 - 7/3	• Linear Regression (Notes L12) • Regularization (Notes L13)	• **A2 due 7/3** • A3 out 7/3	Q9 - L12		• Simple Linear Regression in Matrix Format; • Adding Noise to Regression Predictors;
8	7/6 - 7/10	• Regularization - Contd (Notes L13) • Project Practical Advice • NB & Logistic Reg (Notes L14)		Q10 - L13, 14	• Project midpoint report 7/10 (**Note:** You may submit it without penalty until 7/13.) • Peer Evaluation 7/17	
9	7/13 - 7/17	• Neural Networks (Notes L15) • CNN (Notes L16)	• **A3 due 7/17** • A4 out 7/17	Q11 - L15, 16		• NN Playground • Interactive NN initialization; • The role of a hidden layer; • Back propagation numerical example; • More detailed introduction; • CNN Live Demo; • A guide to an efficient way to build CNN and optimize its hyper-parameters; • Back Propagation in CNN; • Transfer learning in CNN; • Project Scoring Guidance;
10	7/20 - 7/24	• DT (Notes, L17) • RF (Notes, L17) • SVM (Notes, L18)		Q12 - L17 | Q13 - L18		

Week	Dates	Topics	Homework	Quizzes	Project	Readings
11	7/27 - 7/28	• SVM-Kernel (Notes, L19) • Final Instructional Day	• **A4 due 7/27** can submit till 7/31 without any penalty		• Final Project due 7/28 • Peer Evaluation due 7/28	• KKT and SVM • Final Class days (July 28 and 29)

General

General

Attendance

Our class will be offered on campus for both Undergrad (4641) and Grad (7641). Lectures might be recorded IF class has the recording system. Any class that I am able to record [which sometimes does not work even if we have the recording system in place], I will make it available to all students (both undergrad and grad) by the end of the day. The attendance is required for both undergrad and grad. Having students in the class helps me and my students A LOT work with each other for a better environment to facilitate learning. Trust me it will be fun and you will give me a lot of energy to teach better. The fact that you need to listen to the lectures without fast-forwarding me can help you learn the materials much better and you will have the chance to ask questions if you are confused anywhere in the lectures. Also, **the class attendance will be counted toward your class participation at the end of semester.**

Class Deliverables

All class deliverables will be handled via Gradescope except quizzes which will be on Canvas. The time span offered to complete the course objectives is plentiful and deadlines will not be extended under any circumstances. To ensure the class is fair for all students, you will receive zero credit for work submitted after the deadline. Regrade requests should be submitted directly on Gradescope within a defined period after grade publication (we will inform you on that; we only provide a 3 day for the regrade request). Should you find yourself in an impasse with the TA responsible for your grading, feel free to contact the head TA or course instructor on Edstem.

Edstem

Edstem will be the main and only place for the course discussions and announcements. If you have questions, please ask it on Edstem first because 1)

other students may have the same question; 2) you will get help much faster.

- For public homework specific questions, PLEASE use the appropriate TA created mega threads instead of creating a new individual thread.
- If it's something you do not like to discuss publicly on Edstem, you can create a private post on Edstem.
- If the issue you want to discuss is sensitive or you are not comfortable discussing with the whole teaching team in a private post, please create a private post asking us to create a private chat. Either Mahdi or one of the Head TAs will create a private chat for you.
- **Edstem GOOD questions:**
  - "I don't understand this part of the lecture, can you explain it to me?"
  - "This certain part of the hw is not clear to me, would it be possible to explain that more?"
  - "I have a question about the project …"
  - "I found an issue on the website, hw or the lectures, can you clarify …"
  - "Any feedback, suggestions, … would be greatly appreciated."
  - Historically, most of the questions were good!
- **Edstem BAD questions**:
  - "Can you debug my code?"
  - "Can you find where the problem is in my code?"
  - Our team will not do that. You need to be specific about your question

Exceptional Circumstances

Any request for exceptions to these policies should be made in advance when at all possible. Requests should be due to incapacitating illness, personal emergencies, or similarly serious events. Your request **MUST** be accompanied by a supporting letter issued by the Dean of Students before contacting us. Once you acquired the letter, please go this Ed Discussion post and fill out the form and ping us on Ed Discussion using a private post that you filled out the form.

The supporting letter from the Dean of Students should be delivered within a week of the assignment due date in order for us to issue accommodations. If the assignment was due over a week ago, we won't be able to issue accommodations for the assignment.

AI-Based Assistance

We are using the AI assistant policy developed by David Joyner and shared by other classes at Georgia Tech (CS 7643 Deep Learning). The summary is that you

should treat your AI source like a human source, with all accompanying plagiarism implications:

We treat AI-based assistance, such as ChatGPT and Copilot, the same way we treat collaboration with other people: you are welcome to talk about your ideas and work with other people, both inside and outside the class, as well as with AI-based assistants.

However, all work you submit must be your own. You should never include in your assignment anything that was not written directly by you without proper citation (including quotation marks and in-line citation for direct quotes).

Including anything you did not write in your assignment without proper citation will be treated as an academic misconduct case. If you are unsure where the line is between collaborating with AI and copying AI, we recommend the following heuristics:

Heuristic 1: Never hit "Copy" within your conversation with an AI assistant. You can copy your own work into your own conversation, but do not copy anything from the conversation back into your assignment.

Instead, use your interaction with the AI assistant as a learning experience, then let your assignment reflect your improved understanding.

Heuristic 2: Do not have your assignment and the AI agent open at the same time. Similar to the above, use your conversation with the AI as a learning experience, then close the interaction down, open your assignment, and let your assignment reflect your revised knowledge.

This heuristic includes avoiding using AI directly integrated into your composition environment: just as you should not let a classmate write content or code directly into your submission, so also you should avoid using tools that directly add content to your submission.

Deviating from these heuristics does not automatically qualify as academic misconduct; however, following these heuristics essentially guarantees your collaboration will not cross the line into misconduct.

Accommodations for Students with Disabilities

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services (404-894-2563) as soon as possible to

make an appointment to discuss your special needs and to obtain an accommodations letter.

Academic Integrity

- All learners are expected to know and abide by the Georgia Tech Academic Honor Code and the student Code of Conduct.
- Ethical behavior is extremely important in all facets of life.

> 1. Plagiarism is a serious offense. You are responsible for completing your own work. You are not allowed to copy and paste, or paraphrase, or submit materials created or published by others, as if you created the materials. All materials submitted must be your own.
> 2. You may discuss high-level ideas with other students at the "whiteboard" level (e.g., how cross validation works, use hashmap instead of array) and review any relevant materials online. However, each student must write up and submit his or her own answers.
> 3. You must not put your code on public domain (e.g., public GitHub), because a (future) student could copy your code. That student obviously violates the honor code, and you may also be implicated.
> 4. All incidents of suspected dishonesty, plagiarism, or violations of the Georgia Tech Honor Code will be subject to the institute's Academic Integrity procedures (e.g., reported to and directly handled by the Office of Student Integrity (OSI)). Consequences can be severe, e.g., academic probation or dismissal, grade penalties, a 0 grade for assignments concerned, and prohibition from withdrawing from the class.

GUIDELINES

Office Hours

Office Hours

Overview

Office hours will be in a hybrid mode for both online and in-person. We will send an announcement on Ed regarding office hours and when it will start. Please follow the instruction on the Excel Sheet provided on Ed discussion to signup for a slot with one of the TAs. You need to add your name and question of interest. If you require more minutes than the allocated one, please advise the TAs.

Be aware that some TAs have physical hours while others have virtual hours. If you want virtual hours, try to schedule with a corresponding virtual TA. Please do not change the other part of the Excel Sheet.

The TA meetings are designed to be one-on-one. Please do not join another student's meeting. The sole exception to this policy being discussions about the project, in which your fellow team members can also join.

In-person office hours location will be updated on the OH Excel Sheet for each TA.

Rules and Guidelines

There are two types of slots in the Office Hour Spreadsheet: **reserved slots** and **waitlist slots**

- **Reserved Slots**: Students are allowed to hold ONE pending reserved time slot at any time.
  - Let's say it's Tuesday night. Student A signs up for a Wednesday OH slot from 10:00–10:15 AM. Now, student A may NOT put their name on any other reserved time slot. Once the 10:00–10:15 AM OH session has finished, then student A may sign up for another available reserved OH time slot.

    - There is no limit to how many OH sessions a student can attend, but we require you to hold only one active/pending reserved time slot at a time.
- **Waitlist slots**: You are allowed to sign up for multiple waitlist slots per day, but you can only sign up for one slot per TA session.
    - At the start of the TAs OH, if there are regular OH slots available and students on the waitlist, the TA may bump up the waitlisted student to one of the open reserved OH slots.
    - If there are no reserved slots available, a TA may assign an estimated time slot or take the student based on their availability. It is possible that a TA cannot get to any or all of the students on a waitlist.

We have these rules in place so that students can get additional OH help if needed and also to allow availability to a larger subset of students once OH gets busier closer to HW deadlines.

GRADING

Categories

Categories

Exceptional Circumstances

Explanation in General section.

Assignments (40%)

There will be four assignments. Each one is designed to improve and test your understanding of the materials. Assignments will have both programming and written analysis components.

We have 4 big assignments in total. The reason we do not call them projects is because our class has a project as well. Consider each assignment as one individual big project. Assignments take time to finish them. YOU NEED TO START WORKING ON ASSIGNMENTS AS SOON AS THEY ARE OUT. Visit this course's Canvas and GradeScope for the assignment documents. See the schedule table above for deliverable due dates. (Topics are subject to change):

- [10%] HW1: Linear Algebra, Probability and Statistics, Maximum Likelihood Estimation, Optimization, Information Theory
- [10%] HW2: KMeans, Expectation Maximization, Gaussian Mixture Model, Clustering Evaluation
- [10%] HW3: Singular Value Decomposition, Principal Component Analysis, Linear Regression, Regularization, Naive Bayes
- [10%] HW4: Decision Trees, Random Forest, Support Vector Machine, Neural Networks, CNN

You will need to submit all your assignments using Gradescope. Instructions on how to submit your code and written portions will follow with every assignment. **Handwritten solutions WILL NOT BE ACCEPTED** and you will not receive credit for a handwritten submission.

You are required to use Markdown, Latex (watch the tutorial created by our own team and OverLeaf Latex Example in the Video), or a word processing software to generate your solutions to the written questions. Handwritten solutions WILL NOT BE ACCEPTED. You can easily export your Jupyter Notebook to a Python file and import that to your desired python IDE to debug your code for assignments.

All 4 assignments will have a 48-hour **Penalized** Acceptance Period after the assignment due date where we will accept the assignment with a penalty. Submissions turned in during this period will have its grade reduced by a linear percentage deduction commensurate to how much of the 48-hour penalized acceptance period is used. The penalty is capped to a 20% deduction of the submission grade. **Assignments received after the 48-hour Penalized Acceptance Period will receive zero credit**.

This deduction applies separately to each component of the assignment that is submitted separately on Gradescope. Consider an example where a student submits the coding portion of their assignment on time but submits the written portion 12 hours after the deadline. In this case, the coding portion will not be penalized, and the written portion will be subject to a 5% deduction. The late penalty calculation applies as follows:

where x is your assignment grade, t is the number of hours late rounded up, and x' is your final assignment grade with late penalty.

Bonus Points

In every homework, there will be the opportunity to earn bonus credit that will boost your homework grade. These questions greatly help you with your ML understanding, so we highly recommend you complete them. For every homework, **all students can earn up to 10% bonus towards their homework grade** by answering the bonus questions we have in the homework. **Undergrads may earn an extra 6% bonus** (totaling 16%) for solving optional bonus questions only required for grads. Considering that **each homework is worth 10.00%, that means grads can earn up to 1.00% (10% * 10.00%) and undergrads can earn up to 1.6% (16% * 10.00%) bonus** towards their final course grade for each homework.

The value of **each bonus question will be marked in the homework as x% bonus**, which means if you solve it, you get **x% added to your final homework grade out of 100%**. The homeworks required sections will usually be marked in points rather than percent, so you can't create a one-to-one comparison

between required points and bonus credit. For example, **HW1 has up to 120 points**, and say **you earned 102 points in total**. This gives you a **102/120 * 100% = 85% grade** on the homework. If you also **earned 8% worth of total bonus**, you would **receive a 85% + 8% = 93% final grade on the homework**. You can also calculate how many points your bonus credits warrants you, so 8% * 120 points = 9.6 points.

Honor Code

All students are expected to follow the Georgia Tech Academic Honor Code. Because of the large size of our class, if we observe any (even small) similarity/plagiarisms detected by GradeScope or our TAs, **WE WILL DIRECTLY REPORT ALL CASES TO OSI**, which may unfortunately lead to a very harsh outcome.

You are **NOT allowed to share or discuss ANY assignment code, information or answers with other students**. Edstem is the best place to have discussion regarding assignments and course topics. Discussions can be on a whiteboard level with other students such as high level conceptual questions (i.e. what is independency in Naive Bayes model)

In-Class Assessment on Canvas (10%)

We will hold an **in-class assessment** on Canvas. The exam will focus on conceptual questions, and **all students are required to take it**.

- **Make-up exam policy:** The assessment must be taken at the scheduled time. We do **not** offer make-up exams except under the **limited exemptions** described below
- **Attendance and engagement:** Consistent, active participation in the class and on Ed throughout the semester will significantly boost your chances of earning a high score on the assessment.
- **ODS accommodations (student responsibility):** If you are approved for ODS accommodations, you must contact ODS and complete all arrangements **at least 10 working days before the exam date**. You are responsible for meeting this deadline (you already know the exam date from the course schedule). If you do not coordinate with ODS by this deadline, then you will take the exam **in class**, under standard conditions. Given our class size, we cannot make exceptions or last-minute accommodations for missed ODS coordination deadlines.

Make-up Exam

Accepted exemptions and required documentation (only the following): The only acceptable reasons for a make-up exam are:

- **Institute-approved absences** - must be supported by an **Institute Approved Absence letter**.
- **Medical emergencies** - must be supported by a **Dean of Students letter** documenting the medical reason.

If approved, students will take the make-up exam during a **single fixed make-up sitting** on **March 31 at 1:00 PM**.

Not accepted (no exceptions): Any reason **not** covered by an Institute Approved Absence letter or a Dean of Students medical letter is **not** an acceptable exemption. This includes (but is not limited to) job/internship interviews, travel, conferences, work conflicts, other exams preparation, or personal reasons. Please do **not** contact the course staff to request a make-up exam for non-approved reasons — we will not approve it.

Project (30%)

Check the separate project breakdown page for more information on the components of the semester long project.

Note that Project deliverables do not have any Penalized Acceptance Period (i.e. HWs) or Grace Period. For any due dates, please refer to the class schedule table.

Syllabus quiz (1%)

This quiz will test you on the course deadlines and rules. You can simply obtain 1% if you carefully read all the contents of the website and our class rules. We will ask questions like how many quizzes we have in the class? Which days of the week we have most of our deadlines? Is participation required in the course? Etc.

Quizzes (15%)

The total number of quizzes for the semester is listed on the official class schedule (excluding the syllabus quiz and warm-up quiz 0). The weight of each individual quiz is calculated by dividing the total category weight (15%) by the total number of quizzes administered. **All quizzes are mandatory.** Quizzes are open notes and will be proctored using Honorlock. To take a quiz, you must have a stable broadband internet connection, a webcam, a microphone, a valid photo

ID, and Chrome as your web browser. To ensure a smooth experience, please complete Quiz 0 (Practice Quiz) to familiarize yourself with the platform before your first graded quiz. Be sure to consult the Honorlock Student Guide for important technical requirements. Remember, each quiz must be completed independently, without any assistance from others.

The topic of each quiz will coincide very closely with the content covered in class on that week.

Quizzes will have a duration of seven-minutes for Undergrad students and six-minutes for Grad students. Each quiz will have five multiple choice questions . All quizzes will be released on Fridays weekly at 5:00 pm EST and the deadlines will be on Mondays 23:59 EST. To check deadlines for Quizzes, ensure to check the class schedule table. Any possible changes on quizzes dates will be reflected on our course schedule page. Please make sure to check our class website before taking the quiz.

Quizzes are due Monday. Course staff is not guaranteed to be available during Saturday/Sunday.

Quizzes measure your understanding of the topics and they will be mostly conceptual questions.

Quizzes' answers will be released as soon as all our students take them, including our ODS students. Please do not ask any questions about a quiz that you just take on Edstem before we release the answers.

Quizzes questions are selected randomly from our question bank, which means that students will not receive the same questions for their quiz.

Class Participation (4%)

Edstem has statistics which give us many measurements regarding how much a student has been involved on Edstem's activities such as viewing posts, answering questions, asking questions and so on. We use this to account for your Class Participation score. We also will add class attendance to this score. At the end of the semester, we will define a minimum and maximum number of involvement considering all the students and your grade will be defined based on that.

We will RELEASE the class participation score on the last day of the class when we have all the score for projects, quizzes and assignments. If you ask us what is

my participation score before the last day of the class; we will say we do not know. So please be patient.

GRADING

Project Breakdown

Project Breakdown

The project is worth 30% of your grade and will focus on applying machine learning algorithms and methods to real-world datasets to create meaningful insights and useful predictions. You will create a website with GT GitHub Pages for your project which will be used to publish three deliverables: a proposal (5%), a midterm checkpoint (10%), and a final report (15%).

- CS 4641 students are required to use supervised learning. It is highly encouraged to use unsupervised learning methods as well.
- CS 7641 students are required to use both unsupervised and supervised learning.

Team Composition

Each project must be completed in a team of 5. **Students from CS 4641 can only team up with students in CS 4641, and students from CS 7641 can only team up with students in CS 7641.** In addition, each student can only join one team, and it is NOT allowed to be on two teams at the same time.

You will be forming your team on your own. Please use EdStem to find team members. If you cannot find a team, you will be randomly assigned to a team. Note that if you form a team with less than 5 people, we may randomly assign extra teammates.

Once teams have been formed, you will be assigned a mentor. **It is highly recommended to set up meetings with your mentor.**

Peer Evaluation

After the proposal, midterm checkpoint, and final report, students will be required to complete a peer feedback form on CATME for every teammates. **Failure to complete this feedback will result in a grade reduction**.

This measure intends to identify non-contributing members. The teaching team reserves the right to reduce project grades down to a 0 in extreme situations if sufficient evidence is gathered.

General Project Guidance

Please see the proposal, midterm checkpoint, and final report pages for specific guidance and requirements for each project deliverable.

Criteria

Your project will be graded based on the following general criteria:

Was the motivation clear?

- What is the problem?
- Why is it important and why we should care?

Were the dataset and approach used effectively?

- How did you get your dataset?
- What are its characteristics (e.g. number of features, # of records, temporal or not, etc.)?
- Why do you think your approach can effectively solve your problem?
- What is new in your approach?
- Effective visualizations? Are they relevant? Do they help you better understand the project's approaches and ideas?

Were the experiments, results, and conclusion satisfactory?

- What are the results?
- How did you evaluate your approach?
- Why do you think your results are satifactory? If not, what are the limitations? What is work that could be done to improve results?
- How do you compare your method to other methods?
- Effective visualizations? Are they relevant? Do they help you better understand the project's approaches and ideas?

How was effective was the presentation?

- Finished on time?
- Effective visualizations? Are they relevant? Do they help you better understand the project's approaches and ideas?

- Is the use of text concise? Does it describe technical content succinctly? Does the amount of text overwhelm the audience?

How do the results contribute to sustainability in AI? The sustainability in AI is optional and highly recommended and considering them will enhance the quality of the project.

- Do you address any ethical considerations (e.g. bias correction, using fairness metrics to evaluate your model's performance across different groups)?
- Do you address any environmental considerations (e.g. optimizing models for energy efficiency)?
- Do you address any economic considerations (e.g. evaluating the cost-effectiveness and scalability of your solution)?

Helpful Resources

- Sample Projects: On the website under Grading category/Project Breakdown, check Awards Galore for some of the best project examples in our class*.
- Project Ideas: When brainstorming ideas for a project, reference the Stanford Project Examples.
- Dataset: When sourcing datasets, reference the Dataset Ideas.
- Seminars: We will have project seminars where TAs will present their ML projects, prior students' projects, and/or research and industrial projects. Students will also have the opportunity to ask questions about their project. Seminars will be streamed online and recorded. Posts on EdStem will be made for each seminar with details on when and how to join each meeting along with recordings.
- Compute: Google Colaboratory allows free access to run Jupyter Notebooks using GPU resources. The Google Cloud Platform and AWS Educate are also good resources. The GitHub Student Developer Pack also offers free Microsoft Azure and Digital Ocean credits. This semester, we are also offering PACE ICE, Georgia Tech's in-home cluster to students.

GRADING

Proposal

The primary goal of the proposal is to identify a problem that can be solved with machine learning. This includes finding and/or creating a dataset as well as developing a plan for the semester.

Note: As part of research, it is natural that the project may change from the original proposed. Please be sure to document and justify these changes in the midterm and final report.

Proposal Sections & Checklist

1. Introduction/Background: Provide an introduction of your topic and literature review of related work. Briefly explain your dataset and its features, and provide a link to the dataset if possible.

```
✅Literature Review
✅Dataset Description
✅Dataset Link (if applicable)
```

2. Problem Definition: Identify a problem and motivate the need for a solution.

```
✅Problem
✅Motivation
```

3. Methods: Present proposed solutions including specific data processing methods and machine learning algorithms, and elaborate on why you think each will be effective. It is recommended to identify specific functions/classes in existing packages and libraries (i.e. scikit-learn) rather than coding the algorithms from scratch.

```
✅3+ Data Preprocessing Methods Identified
✅3+ ML Algorithms/Models Identified
✅CS 7641: Unsupervised and Supervised Learning Methods Identified
✅CS 4641: Supervised or Unsupervised Learning Methods Identified (S
```

4. (Potential) Results and Discussion: Identify several quantitative metrics you plan to use for the project (i.e. ML Metrics). Present goals in terms of these metrics, and state any expected results.

```
✅3+ Quantitative Metrics
✅Project Goals (recommended to include sustainability and ethical c
✅Expected Results
```

5. References: Cite relevant papers and articles utilizing the IEEE format. All reference in this section must have a matching in-text citation in the body of your proposal text.

```
✅3+ References (preferably peer reviewed)
✅1+ In-Text Citation Per Reference
```

Proposal Submission Requirements

1. Report: The proposal report must be written on a website hosted with GT GitHub Pages. The report must be less than 800 words, and sections (1) through (4) from above count towards this word limit.

In addition to the 5 sections above, please include the following in your report:

- ✅`Gantt Chart`: list each members' planned responsibilities for the entirety of the project. Feel free to use the Fall and Spring semester sample Gantt Chart.
- ✅`Contribution Table`: list all group members' names and explicit contributions in preparing the proposal using the format below.

Name	Proposal Contributions
Member1	Contributions
Member2	Contributions
...	...

2. Video Presentation: The proposal video is a 3-minute recorded presentation that summarizes your proposal using Microsoft PowerPoint, Google Slides, or equivalent. **You must create a YouTube Unlisted video and submit the link.** We will NOT accept submissions which are over 3 minutes or submissions that do not use YouTube Unlisted. Not all group members need to participate in the recording. Some criteria that will be used for evaluation are the following:

- Strong delivery that clearly communicates all sections of the proposal
- Effective slide design which incorporates visuals that are interpretable and legible. Any visuals that are blurry, cut-off, or contain illegible text will not be considered.
- Concise and succinct summary of the proposal and all technical content

3. GitHub Repository: Create a *private* GitHub repository using GT GitHub Enterprise (Please don't use your personal GitHub, just use the GT) all of your teammates and your assigned mentor as a *Collaborator*. There is no need to have any files or code at this point in the repository.

4. Project Award Eligibility: Specify whether you would like to opt-in to be considered for the "Outstanding Project" award. Final project website and contributors of winning projects will be featured on our website! Check the Award Galore

GRADING

Midterm Checkpoint

This is a checkpoint to make sure that you have had major progress in your project. By this point, at least one machine learning model should be implemented and evaluated. Results should be presented using visualizations and quantitative metrics. You will add information to your proposal to create your midterm report.

Note: As part of research, it is natural that the project may change from the original proposed. Please be sure to document and justify any such changes in this report.

Midterm Sections & Checklist

1. Introduction/Background: refer to guidance from the proposal checklist
2. Problem Definition: refer to guidance from the proposal checklist
3. Methods: Present at least one implemented solution including specific data processing method(s) and machine learning algorithm(s). Explain why your chosen model(s) and method(s) were selected.

```
✅1+ Data Preprocessing Method Implemented
✅1+ ML Algorithms/Models Implemented
✅CS 7641: Unsupervised and Supervised Learning Method Implemented
✅CS 4641: Supervised Learning or Unsupervised Method Implemented (S
```

4. Results and Discussion: Discuss the results of your methods and present visualizations and quantitative scoring metrics. What does your visualization/metric tell you? Why did your model perform well/poorly? What are the next steps you plan to take?

```
✅Visualizations
✅Quantitative Metrics
✅Analysis of 1+ Algorithm/Model
✅Next Steps
```

5. References: refer to guidance from the proposal checklist

Midterm Submission Requirements

1. Report: The midterm report must be written on a website hosted with GT GitHub Pages. There is no word limitation for this deliverable. Reuse your proposal and make necessary updates and/or additions.

In addition to the 5 sections above, please include the following in your report:

- ✅`Gantt Chart`: list each members' planned responsibilities for the entirety of the project. Feel free to use the Fall and Spring semester sample Gantt Chart.
- ✅`Contribution Table`: list all group members' names and explicit contributions in preparing the midterm using the format below.

Name	Midterm Contributions
Member1	Contributions
Member2	Contributions
...	...

2. GitHub Repository: Reuse the GitHub repository from the proposal, and add all relevant directories, files, and code. Include a `README.md` file explaining all relevant directories and files using the format below.

`/dir/`: Description of the directory
`/dir/file.txt`: Description of the file

GRADING

Final Report

The final report should present and compare at least 3 models that you have developed throughout the semester. Results should be presented using visualizations and quantitative metrics with a heavy focus on comparing the performance and tradeoffs of each approach. You will add information to your midterm to create your final report.

Note: As part of research, it is natural that the project may change from the original proposed. Please be sure to document and justify any such changes in this report.

Final Sections & Checklist

1. Introduction/Background: refer to guidance from the proposal checklist
2. Problem Definition: refer to guidance from the proposal checklist
3. Methods: Present multiple solutions including specific data processing methods and machine learning algorithms. Explain why your chosen models and methods were selected.

```
✅1+ Data Preprocessing Methods
✅3+ Algorithms/Models
✅CS 7641: Unsupervised and Supervised Learning Method Implemented
✅CS 4641: Supervised or Unsupervised Learning Method Implemented
```

4. Results and Discussion: Discuss the results of your methods. Present visualizations and quantitative scoring metrics to analyze the performance of your models. Compare each approach and explain the tradeoffs, strengths, and limitations of each approach. What does your visualization/metric tell you? Why did each model perform well/poorly? How do the models compare to each other? What are the next steps if any?

```
✅Visualizations
✅Quantitative Metrics
✅Analysis of 3+ Algorithms/Models
✅Comparison of 3+ Algorithms/Models
✅Next Steps
```

5. References: refer to guidance from the proposal checklist

Final Submission Requirements

1. Report: The final report must be written on a website hosted with GT GitHub Pages. There is no word limitation for this deliverable. Reuse your midterm report and make the necessary updates and/or additions.

In addition to the 5 sections above, please include the following in your report:

- ✅`Gantt Chart`: list each members' planned responsibilities for the entirety of the project. Feel free to use the Fall and Spring semester sample Gantt Chart.
- ✅`Contribution Table`: list all group members' names and explicit contributions in preparing the final using the format below.

Name	Final Contributions
Member1	Contributions
Member2	Contributions
...	...

2. Hybrid Presentation: The final presentation is a 9-minute in-person or virtual session, followed by a 3 minute Q&A, that summarizes your project. You will coordinate scheduling with your project mentor, who will oversee presentations from all teams under their guidance. You are encouraged to use Microsoft PowerPoint, Google Slides, or similar tools. Clean, high-quality visualizations play a significant role in the mentor's grading. You may also create an interactive visualization or web application to enhance your presentation. All group members are required to participate.

Some key criteria for evaluation include:

- **Introduction**

    - Clearly explains the problem statement, data, and its significance in the ML context
- **Technical Content & Understanding**
    - Presents model architecture and hyperparameters with clear justification and why those models are chosen.
    - Accurately discusses evaluation metrics and their appropriateness for the problem
- **Results & Analysis**
    - Demonstrates proper interpretation of evaluation metrics
    - Provides meaningful comparisons between different approaches or model iterations
- **Presentation & Communication**
    - Be clear and concise
    - Manage presentation time effectively if multiple team members are presenting (multiple team presentation is encouraged)
    - Use professional and effective visualizations (interactive demos are encouraged)
    - Respond thoughtfully to questions, showcasing deep project knowledge
    - Engage with other teams' projects, offering thoughtful questions or feedback

3. GitHub Repository: Reuse the GitHub repository from the midterm, and add all relevant directories, files, and code. Update the `README.md` file explaining all relevant directories and files using the format below.

`/dir/` : Description of the directory
`/dir/file.txt` : Description of the file

GRADING

Awards Galore

Why Awards Matter

Awards serve as a tangible recognition of the dedication and creativity students bring to their machine learning projects. Each semester, these select honors are chosen **from among hundreds of submitted projects**, making the competition fierce and underscoring the ingenuity and motivation of participating teams. By highlighting outstanding work, we encourage students to push boundaries, refine technical skills, and communicate findings more effectively.

This recognition also benefits students' academic and professional growth—award recipients can feature their projects on resumes and link directly to showcased work for added visibility. Additionally, **all projects remain live** on the class website for future semesters, ensuring that both current and prospective students can discover, learn from, and be inspired by past successes. This tradition fosters continuity in the course and builds a growing body of shared knowledge. Being featured on the website is thus both an honor and a motivator, inspiring future cohorts to strive for the same high standard of excellence.

Project Awardees

Below you will find awardees for each semester. The most recent semester is expanded by default, while previous semesters are collapsed for a cleaner browsing experience. Click on a specific semester to view its award recipients.

▶ **Project Awardees for Spring 2026**

▼ **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 Polisetti, 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

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

Required Course Materials

No textbook will be required for this course, however you are strongly encouraged to complete the readings indicated for each class. You may also find the following books very helpful:

- Learning from data, by Yaser S. Abu-Mostafa
- Pattern recognition and machine learning, by Christopher Bishop
- Machine learning, by Tom Mitchell
- Data Mining: Concepts and Techniques, by Jiawei Han, Micheline Kamber, and Jian Pei
- The Elements of Statistical Learning, by Trevor Hastie, Robert Tibshirani, and Jerome Friedman
- Deep Learning, by Ian Goodfellow, Yoshua Bengio, and Aaron Courville

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

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
- Google public datasets
- Kaggle public datasets
- Awesome Public Datasets
- NYC Taxi data for 2013 Trip Data (11.0GB). 2013 Fare Data (7.7GB). Visualization for a days trip.
- Large datasets publicly available.
- Georgia Tech's campus data (has APIs): bus info, directory, building, T-square, room reservation, building facilities usage (e.g., electricity, lights, A/C, etc.), Oscar/course info/registration, etc.
- Yahoo WebScope
- Data.gov: U.S. Government's open data
- IPEDS data: Postsecondary education data from National Centre for Education Statistics
- Bureau of Labor Statistics data
- Uber data: Anonymized data from over 2 billion trips
- Freebase
- Yelp
- Microsoft Academic Graph
- Numerous APIs from Google (e.g., Maps, Freebase, YouTube, etc.)
- Zillow: real estate listing site
- Numerous graph datasets (large and small): SNAP, Konect
- Movies data: IMDB
- List of lists of datasets for recommendations.
- Million song dataset by Echo Nest. It contains not only the basic information of songs (artist, genre, year, length etc), but also some musical features(like tempo, pitch, key, brightness).
- Dataset about soccer games, players, clubs. No API, but easy to scrape. For a soccer player: transfer history, performance, nationality, birth date, etc. For a soccer club: performance, squad, etc.
- The Free 'Big Data' Sources Everyone Should Know
- Quandl – a dataset search engine for time-series data.
- UCI also has a collection of links to various datasets sorted for various tasks (Classification, Regression, etc)
- Amazon AWS Public Data Sets
- KDD Cup: annual competition in data mining, like Kaggle
- Academic domain: Microsoft Academic Search, DBLP
- Retrosheet: MLB statistics (Game/Play logs)
- Classification datasets

- Various geophysical datasets for the oceans (magnetism, gravity, seismology, etc).
- Social trends
- Beer data Website offline 😳. Older version at web.archive.org
- Academic torrents (terabytes)
- Article Search API from the New York Times (all the way back to 1851!)
- Civil Engineering Dataset
- (Kayak: flight, hotel, car, etc.)
- 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.