Applied Machine Learning/ Machine Learning in Practice

11-344, 11-663, 05-834, 05-434

Course Syllabus
Fall 2025

Draft of Saturday 23rd August, 2025 at 20:12

Instructor

- Kemal Oflazer (ko@andrew.cmu.edu; https://andrew.cmu.edu/user/ko)

Teaching Assistants

- Lawanya Baghel (lbaghel@andrew.cmu.edu)
- Anushka Bhave (abhave@andrew.cmu.edu)
- Charis Graham (csgraham@andrew.cmu.edu)
- Anuj Gupta (anujg2@andrew.cmu.edu)
- Ananya Sane(asane@andrew.cmu.edu)
- Smit Patel (smitp@andrew.cmu.edu)
- Nita Thaveesittikullarp (kthavees@andrew.cmu.edu)

Office Hours

- Kemal Oflazer – Tuesdays and Wednesdays 12-1 pm in-person in GHC 5511 or on Zoom at https://cmu.zoom.us/my/oflazer. Otherwise by appointment – please send an email with several slot options.
- TA office hours and modalities (Zoom/in-person) will be announced later.

Course Websites

- **Canvas** at https://canvas.cmu.edu/courses/49552 (for Module, Additional Reading Handouts, Assignments, Quizzes and Exams)
- **Piazza** at https://piazza.com/cmu/fall2025/0543405834113441166 3/home (for course-related Q/A and announcements– also available from within Canvas; Students are automatically registered to Piazza via Canvas)

- **Calendar of course events on Google Calendar**

Miscellaneous Information

The lectures will be held on Tuesdays and Thursday between 9:30 am and 11:50 am Eastern Time in Wean Hall 7500 Auditorium. Students will be expected to attend lectures in person on a regular basis unless they are sick or quarantined.

Course readings will be from the **required** textbook *Witten, I. H., Frank, E., Hall, M., and Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques,* ***fourth edition****, Elsevier: San Francisco.* You can buy the book as an e-book on `amazon.com`. All readings for the first week are posted as pdfs on Canvas, to give you time to prepare.

This course is cross-listed in HCII and LTI. It is 12 units at all levels (PhD/Master's/Undergrad level).

Some Python programming experience is desirable, but not necessary.

IMPORTANT: This document is not a write-only document (:-). You should read it and understand it fully, and be responsible for it. We may update this document occasionally if any issues or problems are noticed.

Course Description

Machine Learning is concerned with computer programs that enable the behavior of a computer to be learned from examples or experience rather than dictated through rules written by hand. It has practical value in many application areas of computer science such as on-line communities and digital libraries. This class is meant to teach the practical side of machine learning for applications, such as mining newsgroup data or building adaptive user interfaces. There will be a significant project focus, and when you have completed the course, you should be fully prepared to attack new problems using machine learning.

Many students who have taken the course in previous semesters have reported that it was a key factor in getting a competitive internship or job and that it prepared them well for these important next steps in their career.

This course does not assume any prior exposure to machine learning theory or practice. And students from all over campus have been successful in completing this course. However, it should be known that this is a challenging course and will require the full 12 hours per week warranted by a 12 unit course, between readings, quizzes, homework, lectures, and project work.

Grades will be based on assignments (which will be almost weekly) and quizzes, two exams, and a course project. Note that the course project is 40% of your grade, and you cannot pass the course without completing it. Quizzes will be taken online and may be taken as many times as desired before the due date, and the highest grade achieved will be the one that is counted.

In order to take full advantage of opportunities for learning, you will need to do preparation ahead of the scheduled lecture times.

Materials for each week will be posted on Canvas under Modules (organized by weeks). Each module will list the readings, quizzes, and any assignments associated with a week. Copies of slides, any additional readings and other relevant lecture content will also be posted under a module.

To take full advantage of the course, We recommend the following practices:

- By midnight on Monday: do readings for Tuesday and post any questions to Piazza. Slides for the lecture should also be available on Mondays or Tuesday just before the lecture time.
- Participate in Tuesday class unless you are sick. You will almost certainly miss some important information and/or discussions if you do not attend class.
- By midnight on Wednesday: do readings for Thursday, and post any questions to Piazza.
- Participate in Thursday class unless you are sick. You will almost certainly miss some important information and/or discussions if you do not attend class.
- Complete the weekly quiz on Canvas by midnight Thursday. This quiz will be based on the lecture and readings of the week.

Learning Objectives

While it will be essential to learn conceptually how machine learning algorithms work and interact with data, the emphasis will be on effective methodology for using machine learning to solve practical problems. Note that this is not just learning to use a tool bench like Weka. It is about knowing how to conceptualize a problem, knowing how to represent your data, being able to interpret your results properly, doing an effective error analysis, and using the results of the error analysis to make strategic decisions about how to adjust the way you have set up your data and selected and tuned your algorithms.

We will cover a wide range of learning algorithms that can be applied to a variety of problems. In particular, we will cover topics such as decision trees, rule based classification, support vector machines, Bayesian networks, clustering and neural networks. In addition to readings from the course textbook, we will have additional readings from research articles that will be announced ahead of time and distributed on Canvas.

In the last third of the course, there will be *an introductory and non-technical coverage* of the main concepts behind the recent developments in artificial intelligence and machine learning, covering topics such as neural networks and deep learning, word embeddings, large language models and their applications mainly in natural language processing.

Assessments and Grading

Your cumulative numerical grading will be determined by the following assessments:

- Weekly quizzes (10%)
- Assignments (20%)
- Midterm Exams (30%)
  - Midterm 1 on Thursday October 9, 2025 (15%)
  - Midterm 2 on Thursday December 4, 2025 (15%)
- Course Project (40%)

Extra credit activities will be provided for the dual purpose of enhancing your learning experience and/or motivating you to submit your assignments early. A very small percentage of students may earn A+s. A score of more than 100% does not guarantee an A+ grade. A+ grades are only applicable to graduate students and their programs' policies.

Using Canvas

We will use Canvas extensively in this course to distribute important information and for you to take quizzes and exams and turn in assignments and the final project. Be sure your Canvas account is set up so that announcements that get posted are emailed to an account where you will see them and read them. The most useful view of course materials will be found on the

Home tab. There you will find a plan for all the units of the course, including where slides and other materials will be posted over time.

Course-related Q/A will take place mostly over Piazza.

Course Policies

This course is designed to be applicable to a very wide range of student backgrounds. Thus, we have carefully selected a book and set of tools that are usable even by students with no computer programming skills. These tools are adequate if students follow course guidelines on selection of projects/project datasets. However, students with programming skills are welcome to approach the assignments and project with different tools as long as they complete the deliverables.

Assignments will consist of experiments and activities using the Weka toolkit.[1] You may also need to check on the Weka website for unofficial additional plugins for the version of Weka you install.

You should also install the LightSIDE toolkit[2] which will be useful for assignments as well. We will also have some activities in Python,[3] but ample help will be offered to non-programmers. No programming is required on exams or on the course project. You are welcome to do the assignments using packages other than Weka and LightSIDE as long as you turn in the required deliverables as indicated in the assignment instructions.

We recommend highly that you install these tools on your computers very early in the semester to avoid unnecessary delays later when you have to work on assignments.

Assignments should be turned in to Canvas. You should label your submission with your Andrew id and the name of the assignment (e.g., `ko-assignment1.pdf`).

Important Note: Assignments will be assigned on a schedule on Canvas and will be due one week later before midnight, turned into to the corresponding submission link on Canvas.

Important Note: You will receive feedback on your assignments, which is designed to help you learn from your mistakes. Each assignment will be graded out of 5 points. In order to motivate you, your assignment grade will be get a bonus bump if you submit "early":

- If you submit your assignment before or on the due date, we will scale your grade by an additional 10%. So if you did great work and got 5/5, we will give you 5.5/5. If you get 1/5 however, you will get 1.1/5.
- You may submit an assignment up to an additional week later than the due date. In that case, no bonus for you! You will just get the 5/5.
- **Any submissions submitted more than one week after the deadline will get 0/5.**

[1] `https://ml.cms.waikato.ac.nz/weka`

[2] `http://ankara.lti.cs.cmu.edu/side/download.html`

[3] `http://www.python.org/getit/`

Note that assignments are typically one or two weeks apart, so it is up to you to plan how and when you will work on the assignments and submit them.

All assignments have to be submitted by the end of December, 4 2025 – the last days of classes. No assignment submissions after that date will be accepted.

Important Note: Weekly quizzes will need to be completed on Canvas within a certain time frame, typically between 5pm Thursday to 5 pm Friday. Depending on the week, you may be allowed to attempt a quiz multiple times.

Important Note: Exams will serve as formal assessments and will serve to measure your level of understanding of major concepts and their relationships in connection with the course objectives. These will be graded. Details on how exams will be held will be provided in due time.

Course Project Requirements

The term project will involve applying machine learning to a substantial problem of the student's choice. The course project is worth 40% of your grade (5% for your you proposal due after the Fall break, and 35% for the final report due December 12, 2025. **You can not pass this course without doing well on the project.** It will be your responsibility to select a topic for your project and to work on it throughout the semester. A short project proposal will be on the date indicated on the course schedule. You will be required at that time to have obtained your data set and to have run an initial experiment on it.

After the project proposal date, it will not be possible to change your selected data set. Several options are found listed in the Example Projects module on the Home tab on Canvas. Students may select one of these projects or may propose one of their own design. Students who wish to design their own project should check in about their plans with the instructor as early as possible in the semester. See also the Project Advice module for important information about what is required to include in your project report at the end of the semester.

Academic Integrity

All students are expected to be familiar with, and to comply with, the University Policy on Academic Integrity at

`http://www.cmu.edu/policies/student-and-student-life/academic-integrity.html`

Any work submitted as a homework assignment or examination must be entirely your own and may not be derived from the work of others, whether a published or unpublished source, the worldwide web, another student, other textbooks, materials from another course (including prior semesters of this course), or any other person or program. You may not copy, examine, or alter anyone else's homework assignment or computer program, or use a computer program to transcribe or otherwise modify or copy anyone else's files.

Students caught violating academic integrity, will receive an R in the course. Note that the reports students turn in for this course must not be the same one they have done or are doing for a separate course. Turning in a report that was done for a different course constitutes an academic integrity violation.

Note also that using any AI assistance (e.g., ChatGPT) during exams or quizzes is considered an academic integrity violation. However if you find that such tools improve your writing presention and provides a learning opportunity for you, their use would be OK, but caveat.[4]

For Students with Disabilities

Carnegie Mellon University is committed to providing reasonable accommodations for all persons with disabilities. If you have a disability and have an accommodations letter from the Disability Resources office, I encourage you to discuss your accommodations and needs with us as early in the semester as possible. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources (`https://www.cmu.edu/disability-resources/students/`), I encourage you to contact them directly at `access@andrew.cmu.edu`. The Office of Disability Resources will review the information you provide. All information will be considered confidential and only released to appropriate persons on a need to know basis. If you have questions regarding accommodations, you may email the Director of the Office of Disability Resources, Catherine Getchell, at `getchell@cmu.edu`. Further information is also available on Scotty.

If you have approved accommodations that you intend to use for this course, please discuss them with us soon as possible so that we can ensure that you have the support you need.

Student Wellness

It is very important that you take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful. If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support.

Here are some useful contact points:

- Catherine Gretchell, Disability Resources Office at `getchell@cmu.edu`
- Counseling and Psychological Services (CaPS), at 412-268-2922 and/or at `http://www.cmu.edu/counseling/`.

[4] `https://www.media.mit.edu/projects/your-brain-on-chatgpt/overview/`

Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.

Inclusive Excellence Statement

We must treat every individual with respect. We are diverse in many ways, and recognizing this diversity is fundamental to building and maintaining an equitable and inclusive campus community. Diversity can refer to our different perspectives, lived experiences and our identities, including but not limited to race, color, national origin, sex, disability, age, sexual orientation, gender identity, pregnancy or related condition, family status, marital status, parental status, religion, ancestry, veteran status, or genetic information. What makes us different can shape the perspectives our students, faculty, and staff bring to our campus. We, at CMU, will work to promote equity and inclusion because they fuel excellence and innovation, but because we want to pursue justice. We acknowledge our imperfections while we also fully commit to the work, inside and outside of our classrooms, of building and sustaining a campus community that increasingly embraces our core value of inclusion.

Each of us is responsible for creating a safer, more inclusive environment.

Unfortunately, incidents of bias or discrimination do occur, whether intentional or unintentional. They contribute to creating an unwelcoming environment for individuals and groups at the university. Therefore, the university encourages anyone who experiences or observes unfair or hostile treatment on the basis of identity to reach out and **report the concern**. All reports will be reviewed and addressed as appropriate and consistent with our policies.

Tentative List of Lecture Topics

1. Machine Learning Overview, Course Overview
2. Cross-Model and Cross-Technique Concepts
3. Introduction to Model Forms
4. Decision Tree Learning in Weka
5. Probabilities and Weights, Open world assumption
6. More on Naive Bayes Basics
7. Evaluating ML Model Performance; Bayesian Networks
8. Nonlinearity, Support Vector Machines; Kernel Machines
9. Managing Project Data
10. Text features,LightSIDE overview, Introduction to error analysis
11. Parameter tuning, Introduction to cost, Lift factor, Other Charts and Curves
12. Feature Selection, Advanced Parameter Tuning
13. Rule-based Learning
14. Similarity-Clustering
15. Instance-based Learning
16. Advanced Topics in ML
17. Introduction to NLP and Linguistic Representation
18. Language Models
19. Introduction to Neural Networks
20. Word Embeddings
21. Recurrent Networks and Applications; Convolutional Networks
22. Large Language Models: Attention; Encoder and Decoder Models
23. Large Language Models: Decoder Models, Prompting, In-Context Learning, Finetuning,
24. LLM Alignment, Reinforcement Learning with Human Feedback, Instruction-tuning