MACHINE LEARNING AND DATA MINING (PQHS 471)

After completing this machine learning and data mining course, students will be able to explain core machine learning methods, implement algorithms in Python, evaluate models rigorously, collaborate in small groups, and apply these techniques to biomedical and clinical data.

Course Description:

Machine learning and data mining play a central role in modern biomedical, clinical, public health, and social research, where vast and complex datasets require efficient, flexible, and interpretable analytical tools. This course provides a practical and theoretically grounded introduction to key supervised and unsupervised learning methods used to uncover structure in data, build predictive models, and interpret underlying scientific relationships. Students will learn statistical learning principles and a broad set of modeling techniques, including regularization methods (ridge, Lasso), tree-based approaches (bagging, random forests, boosting), support vector machines, neural networks and deep learning, clustering, and other unsupervised learning tools such as principal components analysis. Cross-cutting topics such as model assessment, cross-validation, bootstrap methods, and effective visualization will be emphasized throughout. Applications will draw from biomedical, clinical, and public health research as well as other quantitative disciplines. Students will gain hands-on experience implementing methods in Python, interpreting results, and communicating findings to scientific collaborators. The course combines lectures, demonstrations, computational exercises, homework assignments, and data-driven problem solving tied directly to students' professional interests. Primary material will follow *An Introduction to Statistical Learning with Applications in Python* (James, Witten, Hastie, Tibshirani, and Taylor, 2023), supplemented with additional examples and contemporary developments.

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

1. **Gain proficiency in core supervised learning methods**, including regularization, classification algorithms, tree-based models, support vector machine, and deep learning methods.

2. **Acquire competency in standard and modern unsupervised learning techniques**, including clustering and dimensionality-reduction methods.

3. **Hone practical skills by implementing machine learning methods in Python** and applying them to biomedical and clinical datasets.

4. **Develop the ability to interpret, evaluate, and communicate machine learning results** to collaborators in scientific and clinical research.

PQHS 471 (3 Credits Hours): Course Information

Time and Place:	January 12 – May 7, 2026 **Thursday 2:30 PM – 5 PM** **Wood G-73**
Instructor:	**Abdus Sattar, PhD, LLM**

Office:	Wood building: Wood, G-51
Teaching Assistant:	Towsif Raiyan (email: txr269@case.edu)
Office Hours:	**Mr. Raiyan**: <u>Thursday</u> 1 pm – 2 pm or by appointment **Dr. Sattar**: <u>Monday</u> 1 pm – 2 pm or by appointment
Course Web Page:	sattar.case.edu canvas.case.edu
Text (required):	**An Introduction to Statistical Learning with Applications in Python** *by* James, Witten, Hastie, Tibshirani, and Taylor, 2023

Prerequisites: PQHS 431

- This course is designed for advanced undergraduate students, and graduate students in Biostatistics or other sciences with background and adequate preparation in statistical methods (at least one statistics course, equivalent to PQHS 431 course experience). The course will use statistical formulas (e.g., probability mass functions) and theories (e.g., likelihood theory) in explaining various concepts of machine learning and data mining. Undergraduate students require the instructor's permission for enrollment.

- Knowledge of statistical computing is required. We aim to use Python for all computing. Some programming experience will be helpful.

PQHS 471: Course Requirements & Grading

Assignments:

There will be six (6) homework assignments, one (1) midterm exam, and one (1) final exam. No late assignments will be accepted unless you have a university-excused absence. Real data analysis problems drawn from scientific studies will give you the opportunity to demonstrate your machine learning and data mining knowledge through model development, evaluation, and interpretation.

Grading Scale: The course grade will be determined according to the following,

- Homework   40%
- Midterm exam   30%
- Final exam   30%

Midterm and Final Exams:

Midterm Exam – **March 5, 2026** (test will cover topics until 2/26). Time, 3:30 – 5:00 PM (One and half-hour).

Final Exam – **April 30 – May 7** (Chapters 9 and 10). TBD (Two hours).

PQHS 471: Course contents and Timeline (Tentative)

Week	Date	Topics	Textbook Sections
1	**01/15**	Introduction to machine learning	1
		Basics of Data Mining	2.1
		Matrix algebra preliminaries	Slides
		Lab: Introduction to Python	2.3
2	**01/22**	Data mining: Clustering	12.4
		Lab: Clustering	12.5.3
3	**01/29**	Principal Component Analysis	12.2
		Lab: PCA	12.5.1
	02/02	**Homework 1 due by Feb 2 (Monday)**	
4	**02/05**	Classification, logistic regression, LDA, QDA, KNN, GLM	Ch. 4
		Lab: Supervised learning	4.7
5	**02/12**	Resampling methods (CV, bootstrap)	5.1, 5.2
		Lab: Cross-validation and bootstrapping	5.3
	02/16	**Homework 2 due by Feb 16 (Monday)**	
6	**02/19**	Linear model selection & regularization	6.1, 6.2
		Lab: model selection & regularization	6.5
	02/23	**Homework 3 due by Feb 23 (Monday)**	
7	**02/26**	Basics of tree-based methods, Bagging, random forest, boosting	Ch. 8
		Lab: Tree-based methods	8.3
8	**03/05**	Bayesian adaptive regression trees	8.2.4
		2:30 PM – 3:20 PM (lecture part)	
		Midterm Exam 3:30 – 5:00 pm	
9	**03/12**	**Spring Break (no class)**	
10	**03/19**	Support Vector Machine (SVM)	Ch. 9
		Lab: SVM	9.6

	03/23	**Homework 4 due by March 23 (Monday)**	
11	03/26	Deep learning basics: SLNN, MLNN	10.1, 10.2
		Lab: Deep learning coding in Python	10.9.1, 10.9.2
12	04/02	Convolutional Neural Networks	10.3
		Lab: CNN coding in Python	10.9.3
	04/06	**Homework 5 due by April 06 (Monday)**	
13	04/09	Recurrent Neural Networks	10.5
		Lab: RNN coding in Python	10.9.6
14	04/16	Fitting Neural Networks	
		Lab: NN fitting in Python	
	04/20	**Homework 6 due by April 20 (Monday)**	
		Final Exam: TBD	

POLICIES:

University Enrichment & Engagement. It is the intent that all students regardless of their background and perspective be well-served by this class. Further, we intend to present material whose content is respectful and deliver it in a respectful manner. We expect that all students, instructors and guests will help foster an atmosphere of respect, trust and safety in the classroom.

If you have suggestions for how to make the class content or environment more respectful, or have specific incidents to report, please reach out to the instructor. If you are not comfortable reaching out to the instructor, feel free to reach out to someone else, such as the School of Medicine Graduate Education Office (som-geo@case.edu) or the Office of Campus Enrichment and Engagement. More information on University policy and resources are available on their website.

Disability Accommodations. In accordance with federal law, if you have a documented disability, you may be eligible to request accommodations from Disability Resources. In order to be considered for accommodations you must first register with the Disability Resources office. Please contact their office to register at 216.368.5230 or get more information on how to begin the process. Please keep in mind that accommodations are not retroactive.

Academic Integrity. Visit the following website for the Academic Integrity for undergraduate courses:

https://bulletin.case.edu/undergraduate-academics/academic-integrity/academic-integrity.pdf

Visit the following website for the Academic Integrity for graduate courses:

https://case.edu/gradstudies/sites/default/files/2018-04/SGS-Academic-Integrity-Policies-and-Rules.pdf

AI (Artificial Intelligence). Students at Case Western Reserve University are expected to uphold the highest ethical standards of academic conduct. Academic integrity addresses all forms of academic dishonesty, including cheating, plagiarism, misrepresentation, obstruction, and submitting without permission work to one course that was completed for another course. Please review the complete academic integrity policy for additional information, including an overview of the processes & procedures for reporting academic misconduct.

Restricted AI Usage for Specific Tasks:

- Students are expected to use AI responsibly and critically evaluate the outputs of AI tools, ensuring transparency, accountability, and ethical considerations in their utilization.
- Students can use basic AI tools for specific tasks approved by the instructor, such as grammar checking or citation formatting.
- The use of AI should be justified based on its relevance to the learning objectives, and students must demonstrate a clear understanding of both the AI methods employed and their academic application.
- Students are encouraged to consult with the instructor if they have any questions or concerns regarding the appropriate use of AI in this course.
- Any homework problem solved with the help of AI must be clearly disclosed in your submitted work.

Incomplete grades. **The grade of Incomplete is assigned at the discretion of an instructor provided that:**
1.   There are extenuating circumstances, explained to the instructor before the assignment of the grade, which clearly justify an extension of time beyond the requirements established for other students in the class. It is the student's responsibility to notify the instructor of the circumstances preventing completion.
2.   The student has been passing the course and only a small segment of the course remains to be completed, such as a term paper, for which the extenuating circumstances justify a special exception.

An Incomplete grade may not be assigned if a student is absent from a final examination, unless the dean has authorized the absence. Unauthorized absence from a final examination will result in a failing grade. When the student completes the work, the Incomplete is changed to an A, B, C, D, P, F, or NP.

All work for the incomplete grade must be made up and the change of grade recorded in the Office of the University Registrar by the date specified by the instructor, but no later than the last day of class in the semester following the one in which the Incomplete was received. A student who has a permanent Incomplete for a required course must retake the course in a later term. If the student cannot complete the work by the end of the following semester, he or she must petition for an extension which must be endorsed by the instructor, explain the reasons why the work has not been completed, and include a new date for completion. Students will be allowed only one extension of no more than one additional semester to complete the work for an Incomplete grade.

Laptop and cell phone use. Although having a laptop in class opens up new learning possibilities for students, it can be used in ways that are appropriate. It is easy for your laptop to become a distraction to you and to those around you. Laptops are to be used only when essential to the task at hand such as during the lab portion of the class. Please turn off or silence all cell/smart phones, tablets, and other electronic devices for the duration of the course. Inappropriate uses will be noted and may affect the final grade.

Academic Integrity. Any violation of the University's Code of Ethics will <u>not</u> be tolerated. All forms of academic dishonesty including cheating, plagiarism, misrepresentation, and obstruction are violations of academic integrity standards and will result in a minimum penalty of receiving a zero for the assignment, the potential for failing the entire course. Cheating includes copying from another's work, falsifying problem solutions or laboratory reports, or using unauthorized sources, notes or computer programs. Plagiarism includes the presentation, without proper attribution, of another's words or ideas from printed or electronic sources. It is also plagiarism to submit, without the instructor's consent, an assignment in one class previously submitted in another. Misrepresentation includes forgery of official academic documents, the presentation of altered or falsified documents or testimony to a university office or official, taking an exam for another student, or lying about personal circumstances to postpone tests or assignments. Obstruction occurs when a student engages in unreasonable conduct that interferes with another's ability to conduct scholarly activity. Destroying a student's computer file, stealing a student's notebook, and stealing a book on reserve in the library are examples of obstruction.

In addition, the incident will be reported to the Dean of Undergraduate Studies and Academic Review Board for undergraduates or Senior Associate Dean of Graduate Studies, for Graduate Students. The CWRU Statement of Ethics for graduate students can be found here:

http://case.edu/gradstudies/about-the-school/policies-procedures/

Plagiarism. Unless specifically stated otherwise, <u>the faculty of the PQHS expect and require original writing for all</u> <u>assignments given.</u> Submitting plagiarized work for an academic requirement is a violation of the academic integrity standards set forth by the University. Plagiarism is the representation of another's work or ideas as one's own; it includes the unacknowledged, word-for-word use and/or paraphrasing of another person's work, and/or the inappropriate unacknowledged use of another person's ideas. Submitting substantially the same work to satisfy requirements for one course that has been submitted in satisfaction of requirements for another course, without permission of the instructor of the course for which the work is being submitted, is also prohibited.

Course Audit. Each attendee and participant in a PQHS/MPHP/CRSP course must be registered for the course. The Department of Population and Quantitative Health Sciences does not permit the auditing of courses, either officially or unofficially, without the written approval of the instructor and the Vice Chair for Education for PQHS. Students looking to take a course outside the requirements of their degree program are encouraged to make use of the Fellowship Course option available through the School of Graduate Studies.

Equity and Title IX. Title IX of Education Amendments of 1972 ("Title IX") prohibits discrimination based on sex, including sexual harassment, in education programs that receive federal funding. This includes providing support to anyone impacted by sex discrimination or sexual harassment. The Title IX Coordinator is responsible for implementing Title IX and is in CWRU's Office of Equity. The Title IX Coordinator is also responsible for making sure that the procedures in CWRU's policy prohibiting sex discrimination and sexual harassment are followed and all persons are treated fairly and with respect. Members of the CWRU community and others should direct questions about the application of Title IX to situations impacting the CWRU community, and its educational programs and activities, to the Title IX Coordinator. Information on the Office of Equity and Title IX is located on the Office of Equity webpage.

CWRU Public Safety. The Case Western Reserve University Division of Public Safety provides quality service, safety, and a positive educational experience for the campus community. Members of the CWRU community – students, faculty, staff, and guests – are encouraged to report all criminal actions,

emergencies, or other public safety related incidents to CWRU Division of Public Safety. For an emergency, call 216-368-3333. For a non-emergency, call 216-368-3300. Members of the community are encouraged to download the Spartan Safe app which offers easy access to critical resources and safety tools.

~ End ~