University of Texas at El Paso
School of Sciences
Department of Earth, Environmental and Resource Sciences

GEOL4315 Machine Learning for Scientists and Engineers
GEOL5324/GEOL6324 Geocomputation

Spring 2023

Instructor Information

Instructor: Dr. Hernan A. Moreno
Classroom: Theory GEOL 302 (T) & Lab GEOL 404 (R)
Meetings: TR 3:00 PM – 4:20 PM
Office: Geologic Sciences Building Suite 321 A
Office Hours: TR 11:00 AM – 12:00 PM
Learning Management Site: Blackboard
Teaching Assistant: Jorge Mayo- jamayo@miners.utep.edu
Teaching Assistant Hours: TR 4:20- 5:20 PM at the classroom

Course Description

This course provides knowledge, skills, and tools for data science and machine learning applied to earth, environmental sciences, and engineering. Students are expected to have a minimum background in linear algebra, statistics, and probability. The course covers a selection of supervised learning methods from the most fundamental (K-Nearest Neighbors, Decision Trees, and Linear and Logistic Regression) to more advanced methods (Random Forests, Boosting, Support Vector Machines, Deep Neural Networks) plus commonly unsupervised methods (Generative Modeling, k-means, PCA, Autoencoders and Generative Adversarial Methods). The course also offers fundamental knowledge of useful concepts like loss and cost functions, maximum likelihood, bias-variance decomposition, ensemble averaging, kernels, and Bayesian approaches, along with useful techniques such as regularization, cross-validation, evaluation metrics, and optimization. The final chapters discuss some common issues when tackling a machine learning problem and the ethics when developing learning algorithms. Along with the theoretical part, laboratory exercises and paper discussions will be developed so that students get hands-on experience and see the learned concepts applied to earth science and engineering. Labs will be developed using Python as the main programming language. A final semester project will be developed and presented at the end of the course by each of the students as a way to apply one or more ML techniques to an engineering or earth science problem.

Learning Outcomes/Course Objectives

Students completing this ML/Geocomputation course will be able to:

- Describe the fundamentals of supervised learning, its application, pros, and cons.

- Understand model validation as a crucial step to creating model confidence.
- Create a neural network model to resolve a problem in engineering or geoscience.
- Understand and apply unsupervised learning models.
- Describe the main issues, limitations, and ethical responsibilities of ML and AI models.

Course Materials (Book is optional)

- **Lindholm, A., Wahlström, N., Linsdten, F., and Schön, T. (2022)** Machine Learning: A first course for Engineers and Scientists. Cambridge University Press. 338 pp.

The class materials, including lecture presentations, scientific articles, and laboratory handouts, can be found on the blackboard. Other readings will be provided through Blackboard.

Additional Textbook References (Optional)

- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2013). An Introduction to Statistical Learning: with Applications in R. New York: Springer.
- Geron, Aurelien (2017). Hands-On Machine Learning with Scikit-Learn & TensorFlow: concepts, tools, and techniques to build intelligent systems. Beijing: O`Reilly.
- Andreas C. Müller and Sarah Guido (2016). Introduction to Machine Learning with Python: A Guide for Data Scientists. O`Reilly.

Course Structure

The course format will be lectures, lab assignments, article presentations/discussions, and a long-term project. Students will gain hands-on experience using Python for script writing and execution. However, the knowledge gained during the practical sessions will easily be transferable to other platforms like R, Julia, etc.

Lectures: The lecture section will include more than instructor-led discussions. The instructor expects students to attend classes and only work on course-related materials during that time. To help students gain a better insight into Machine Learning, the lecture session will also include additional time for application demonstrations, in-class exercises, and more. Attendance will be taken during every class and lab meeting, and a total of 2% of the class grade may be earned at the end of the semester to encourage and reward students that always attend all lectures and labs. Participation is essential to the course and will be assessed based on in-class activities. A total of 3% will be awarded to students participating in the in-class questions or paper presentations (at least once during the semester). This active learning strategy is designed to enhance your learning and provide you with an opportunity for reflection, elaboration, and application.

Lab Exercises: The laboratory sessions are an essential component of Machine Learning training since they provide students with hands-on experience in ML coding to consolidate their understanding of theoretical concepts and analytical techniques. In this course, we will be using Jupyter Notebooks (or Jupyter Books through Kaggle) so that students develop their labs in Python language. The Labs include an initial Python Crash Course developed through Coursera Link on Blackboard (details later) and followed by sequential Labs from the Kaggle free online platform. A total of eight (8) mandatory and one (1) extra-credit (5%) labs will be completed. Most lab work coincides with skills and concepts learned in lectures and readings. It is, therefore, vital for students to attend class so that they understand the Lab exercises and complete them on time. You will have time to advance lab assignments during scheduled class meetings (Thursdays), but you will require time outside class to complete them. Lab assignments must be turned in via Blackboard by the due date specified in the lab document, and the submitted assignment must be your own work. You should use your class time devoted to the lab activities wisely and not expect to leave class early. You are welcome to ask the instructor for help outside of class as long as you are making an effort to work while in class. Do not expect to miss lab time and to get extra outside class help to compensate.

Paper Presentation and Discussion: Graduate students will be tasked with presenting a scientific article assigned at the beginning of the semester.
Presentations will be 30 minutes long, with 10 minutes for comments and questions from the audience.

Long-Term Project: Students will complete one research project due at the end of the semester (final version). The project builds on skills, and concepts learned throughout the entire course, but certain components (such as the project "outline" and identified data sources") will be discussed before the Spring Break (i.e., the first phase). Students taking Geocomputation as part of the GIST certificate must select a topic, including geospatial data, to develop their machine learning project. Please plan accordingly and start early, as the project should take some 20-40 hours of your time to develop. More specific details on the project guidelines will be provided in class. Students should expect to spend time outside scheduled hours to complete the course project. Students are expected to develop a clear science question for a paper-style work. More details will be provided soon, but students should start thinking and reading about a topic they would like to research. The final project will assess the student's ability to complete a research project comparable to what a machine learning expert will see working in the field.

Due Dates

Lab Exercises: In this course, there is an absolute due date for completing each lab assignment, which will be indicated on the lab handouts. The reason for this is that each lab conceptually builds on previous labs, so it is important that students do not fall behind schedule. If you do, you may find it hard to catch up. Labs must be submitted on time and to the correct blackboard assignment location to be considered for grading. We also encourage students to start working on labs well in advance of the assigned deadline. Note that we will not make any exceptions to this policy for any student unless there is a documented extenuating circumstance. Please plan ahead using the course schedule at the end of this syllabus and deadlines as posted on Blackboard.

Long-term Project: Each of the two phases of the final project will have an assigned in-class due date. The first phase (starting in March 30th) will be worth 5% of the total course grade. In the second phase (May 11th), students will prepare a scientific manuscript and a presentation compiling the work's objectives, methods, results, and conclusions. Similar to the information above, note that late final project assignments will be accepted for a 30% reduction if turned in one day late; afterward, late final projects will not be accepted for any reason after the assigned due date.

Late and Missing Work: Late laboratory work submissions are accepted, but 20% of the total grade will be deducted each day after an assignment deadline. Make-up assignments or late submissions can only be allowed in the event of a documented medical or family emergency. If you encounter an emergency, you must notify me on or before the day of the assignment or exam due date. Documentation could include a note from a physician, a hospital admittance slip, or correspondence from an academic advisor or the Dean of Students. Foreseeable excused absences, such as participation in university-sanctioned athletic or academic events, require documentation and notifying the Instructor at least one week in advance. In the case of foreseeable absences, you must turn in work early rather than late. In each of these situations, the student must communicate and keep the instructor informed.

Rules and Policies

Notes and Lectures: Any notes, labs, or lecture materials or for personal use only, and their sale or distribution to people outside the class is not permitted.

Blackboard Course Management System and Email Communication: This class uses the Blackboard learning platform that you can access through your my.utep.edu portal (use your e-mail address and password to sign in). For this class, all communication will be via the regular UTEP email system, so avoid using the Blackboard email function. As a requirement of this class, you must regularly check your UTEP email for posts that include announcements and personal messages to students. You are responsible for reading all your emails, so not having seen or read an email containing important course-related announcements will not be an acceptable excuse for failure to complete assignments on time unless there is a verifiable technical issue. Please do not use your private email addresses for course-related communication.

Academic Honesty: Academic honesty is fundamental to the activities and principles of the University. Students are encouraged to share ideas, but you must do your own work (except in the case of collaborative group assignments). For example, cheating, plagiarism, and collusion will not be tolerated. Academic dishonesty has grave consequences that range from probation, expulsion, and failing the course. If you are found to be cheating, plagiarizing, or colluding, you will be subject to disciplinary action per UTEP catalog policy. Refer to http://sa.utep.edu/osccr/academic-integrity/ for further information.

Cell Phone Policy: Cell phones are not permitted for use at any time during this class. Taking calls, texting, taking pictures, or recording video during class is strictly prohibited. Any attempt to use a cell phone during class or lab will be seen as an attempt to cheat and subject to the policies defined in the UTEP Academic Misconduct Code.

Students with Disabilities: UTEP is committed to equity in providing educational services to students with disabilities. If you have (or suspect) a disability that requires accommodations or may affect your performance in this course, tell the professor so that we can work with the Center for Accommodations and Support Services (CASS) to meet your needs. You can contact CASS at (915) 747-5148, or by email to cass@utep.edu, or visit their office located in UTEP Union East, Room 106. For additional information, please visit the CASS website at www.sa.utep.edu/cass.

Help with Writing: The University Writing Center (UWC) (Library Building, 2nd [ground] floor, Room 227 [adjacent to the Collaborative Learning Center], Tel. 747-5112) provides walk-in assistance with writing to all UTEP students at no cost. If you struggle with writing, the tutors at the UWC can help you! Check the website – http://uwc.utep.edu/ for a tutoring schedule.

Religious Absence: It is the policy of the University to excuse the absences of students that result from religious observances and to reschedule examinations and additional required classwork that may fall on religious holidays, without penalty. Students who anticipate being absent from class due to the observation of major religious activity must provide written notice of the dates to the instructor by the second week of the semester.

COVID-19 PRECAUTION STATEMENT

Please stay home if you have been diagnosed with COVID-19 or are experiencing COVID-19 symptoms. If you feel unwell, please let me know as soon as possible so we can work on appropriate accommodations. If you have tested positive for COVID-19, you are encouraged to report your results to covidaction@utep.edu, so that the Dean of Students Office can provide you with support and help with communication with your professors. The Student Health Center is equipped to provide COVID-19 testing.

The Center for Disease Control and Prevention recommends that people in areas of substantial or high COVID-19 transmission wear face masks indoors in groups of people. The best way that Miners can take care of Miners is to get the vaccine. If you still need the vaccine, it is widely available in the El Paso area and will be available at no charge on campus during the first week of classes. For more information about the current rates, testing, and vaccinations, please visit epstrong.org.

Grading

Description	No	Graduates	Undergraduates
Laboratory Assignments	8	40% (5% each)	80% (10% each)
Paper Presentations	1	15%	-
Long Term Project	1	40% (5%+35%)	15% (5%+10%)
Class attendance + participation	1	2%+3%	2%+3%

Final grade table

Percent grades will be rounded to one decimal place, and letter grades will have the following equivalence. Note, beyond this table, we will not round grades anymore for any reason, so please do not ask. It is not fair to students who legitimately earned the same grade. Please use the extra-credit opportunities given throughout the course to improve your grades:

Letter Grade	Grade Point	Percentage
A	4.0	89.5 to 100
B	3.3	79.5 to 89.4
C	2.0	69.5 to 79.4
D	1.0	59.5 to 69.4
F	0.0	59.4 to 0

Course syllabus

Week	Date	Topic	Reading	Due
1	17 Jan (T)	**Introduction, syllabus and logistics.** **Lab#0. Python Installation** **Lab#1. Python Crash Course** **Lecture #1. Machine Learning Exemplified**	Syllabus Ch 1 Lindholm Google Certificate	Lab 1 due 02/01 @5 PM
	19 Jan (R)	**Lecture #1. Machine Learning Exemplified** **Lab#1. Python Crash Course**	Ch 1 Lindholm Google Certificate	
2	24 Jan (T)	**Lecture #2. Supervised Learning:** K-Nearest Neighbors (KNN) **Lab#1. Python Crash Course**	Ch 2 Lindholm Google Certificate	
	26 Jan (R)	**Lecture #3. Supervised Learning:** Decision trees. **Lab #1 (Extra-credit 5%):** mlcourse.ai on Kaggle – Pre-requisites, Topic1 and Topic 2.	Ch 2 Lindholm Lab #1 Extra-credit	Lab 1 Extra-Credit due 02/08
3	31 Jan (T)	**Lecture #4. Parametric Models and Statistical Learning:** Linear regression, classification and logistic regression	Ch 3 Lindholm	
	2 Feb (R)	**Lab #2:** Decision Trees and K-NN	mlcourse.ai on Kaggle	Lab 2 due 02/08 @5 PM
4	7 Feb (T)	**Lecture #5. Parametric Models and Statistical Learning:** Polynomial regression and regularisation, generalized linear models.	Ch 3 Lindholm	Paper selection @ 5 PM

Week	Date	Topic	Reading	Due
	9 Feb (R)	**Lab #3:** Logistic Regression	mlcourse.ai on Kaggle	Lab 3 due 02/15 @5 PM
5	14 Feb (T)	**Lecture #6. Understanding, evaluating and improving performance:** Expected new data error and estimation (E_{new}). Training error - generalization gap decomposition. The bias-variance decomposition of E_{new}. Evaluating binary classifiers.	Ch 4 Lindholm	
	16 Feb (R)	**Lab #4:** Validation and learning curves	mlcourse.ai on Kaggle	Lab 4 due 02/27 @5 PM
6	21 Feb (T)	**Lecture #7. Learning parametric models and optimization:** Principles of parametric modeling, loss functions and likelihood models.	Ch 5 Lindholm	
	23 Feb (R)	**Lecture #8. Learning parametric models and optimization:** Regularisation, parameter optimization, hyper-parameter optimization.	Ch 5 Lindholm	
7	28 Feb (T)	**Lecture #9. Ensemble methods, bagging and boosting:** Bagging, random forest	Ch 7 Lindholm	
	2 Mar (R)	**Lab #5:** Exploring OLS, Lasso and Random Forest in a regression task	mlcourse.ai on Kaggle	Lab 5 due 03/08 @5 PM
8	7 Mar (T)	**Lecture #10. Ensemble methods, bagging and boosting:** Boosting and AdaBoost, gradient boosting.	Ch 7 Lindholm	
	9 Mar (R)	**Lab #6:** Gradient Boosting	mlcourse.ai on Kaggle	Lab 6 due 03/29 @5 PM
9	14 Mar (T)	**No class – Spring Break**		
	16 Mar (R)	**No class – Spring Break**		
10	21 Mar (T)	**Paper Presentation #1** Random Forest, Gradient Boost	Selected PDF Papers	
	23 Mar (R)	**Lecture #11. Neural Networks and Deep Learning:** The ANN model, training Convolutional neural networks, dropout.	Ch 6 Lindholm	
11	28 Mar (T)	**Lecture #12. Neural Networks and Deep Learning:** The ANN model, training Convolutional neural networks, dropout.	Ch 6 Lindholm	Project Proposal @ 5 PM
	30 Mar (R)	**Lab #7:** ANN and CNN	Deep NN on Kaggle	Lab 7 due 04/19 @5 PM
12	4 April (T)	**Paper Presentation #2:** ANN, CNN	Selected PDF Papers	
	6 April (R)	**Lecture #13. Non-linear input transformations and kernels:** Creating features by N-L input transformations, kernel ridge regression, support vector regression, kernel theory, support vector classification.	Ch 8 Lindholm	
13	11 April (T)	**Paper Presentation #3:** Kernel-Rigdge Regression, Support Vector Machines	--	
	13 April (R)	**Lecture #14. Unsupervised learning:** The gaussian mixture model and discriminant analysis	Ch 10 Lindholm	

Week	Date	Topic	Reading	Due
14	18 April (T)	**Lecture #15. Unsupervised learning:** Deep generative models, representation learning and dimensionality reduction.	Ch 10 Lindholm	
	20 April (R)	**Lab # 8:** PCA and Clustering	mlcourse.ai on Kaggle	Lab 8 due 05/03 @5 PM
15	25 April (T)	**Paper Presentation #4:** LDA and QDA, Cluster analysis	Selected PDF Papers	
	27 April (R)	**Paper Presentation #5:** GANs, Autoencoders	Selected PDF Papers	
16	2 May (T)	**Paper Presentation #6:** Principal component analysis, discovering physical concepts with NNs	Selected PDF Papers	
	4 May (R)	**Lecture #16. Ethics in ML:** Fairness and error functions, misleading claims on performance, limitations of training data.	Ch 12 Lindholm	
17	9 May (T)	Finals Week – no class		
	11 May (R)	**Long Term Report Due and Poster Presentation** (4:00 PM – 6:45 PM) https://www.utep.edu/student-affairs/registrar/Scheduling/Final%20Exam%20Schedule%20Spring%202023.pdf		Paper due 05/11 @1 PM

Course Resources: Where you can go for assistance

UTEP provides a variety of student services and support:
Technology Resources

- Help Desk: Students experiencing technological challenges (email, Blackboard, software, etc.) can submit a ticket to the UTEP Helpdesk for assistance. Contact the Helpdesk via phone, email, chat, website, or in person if on campus.

Academic Resources

- UTEP Library: Access a wide range of resources including online, full-text access to thousands of journals and eBooks plus reference service and librarian assistance for enrolled students.
- University Writing Center (UWC): Submit papers here for assistance with writing style and formatting, ask a tutor for help and explore other writing resources.
- Math Tutoring Center (MaRCS): Ask a tutor for help and explore other available math resources.
- History Tutoring Center (HTC): Receive assistance with writing history papers, get help from a tutor and explore other history resources.
- RefWorks: A bibliographic citation tool; check out the RefWorks tutorial and Fact Sheet and Quick-Start Guide.

Individual Resources

- Military Student Success Center: Assists personnel in any branch of service to reach their educational goals.

- Counseling and Psychological Services: Provides a variety of counseling services including individual, couples, and group sessions as well as career and disability assessments.