



SECTION SYLLABUS

GRADTDA 5620

GRADTDA 5620 – Practical Learning and Mining for Big Data
Summer 2024, Online

COURSE OVERVIEW

Instructor

CSE Instructor: Ali AlilooeeDolatabad

Email address: alilooeeDolatabad.1@osu.edu

Office hours: TBD

Statistics Instructor: Thomas Metzger

Email address: metzger.181@osu.edu; **please do not use my Buckeyemail address.**

Office hours: Mondays, 5-6 pm or by appointment

Course Meeting Information

Meeting Days: Optional but encouraged live sessions held Mondays, 6-7 pm

Prerecorded lectures posted to Carmen

Location: Via Zoom

Course description

Building on computational and interpretative skills in data analytics and computing foundations, students will explore practical ways to create data mining and machine learning workflows. Students will learn to mine associations and patterns, to classify, and build predictive models and recommendation systems for data and questions in the context of enterprises. Further, students will build both custom and public domain tools to infer from both machine and deep learning models and make predictions.

Course learning outcomes

The goals for this course are:

1. Students will gain proficiency in the knowledge discovery process; they will be able to formulate questions pertaining to associations and pattern mining and create workflows that can classify and mine data in a scalable manner.
2. Students will understand the statistical learning framework, including core concepts such as loss, learning, and generalization; they will be able to judge when the framework is applicable and formulate contextual problems from their enterprises within this framework.
3. Students will understand the rationale and algorithms behind statistical learning methods, and they will know the relative merits and limitations of these methods. Students will be able to quantitatively evaluate, validate and compare different statistical learning methods within the context of their enterprises and diagnose errors in a machine learning system.
4. Students will be familiar with the construction of workflows with publicly available and custom tools for multitude of enterprise data; build machine learning systems in complex settings including when training and test sets are mismatched and obtain performance comparable to or surpassing human-level performance.
5. Students will become familiar with the latest and emerging techniques of machine learning and deep learning; they will learn to deploy and work optimally with textbook neural networks and emergent deep learning networks.

HOW THIS COURSE WORKS

Mode of delivery: This course is 100% online. It is a mix of synchronous and asynchronous delivery with lectures delivered via asynchronous video and synchronous Zoom Q&A, activity, and discussion sections.

Pace of online activities: This course is divided into **weekly modules** that are released one week ahead of time. Students are expected to keep pace with weekly deadlines but may schedule their efforts freely within that time frame.

Credit hours and work expectations: This is a **3-credit-hour course**. According to [Ohio State policy](#), students should expect around 3 hours per week of time spent on direct instruction (instructor content and Carmen activities, for example) in addition to 6 hours of homework (reading and assignment preparation, for example) to receive a grade of (C) average.

Attendance and participation requirements: Because this is an online course, your attendance is based on your online activity and participation. The following is a summary of everyone's expected participation:

- **Participating in online activities for attendance: AT LEAST ONCE PER WEEK**
You are expected to log in to the course in Carmen every week. (During most weeks you will probably log in many times.) If you have a situation that might cause you to miss multiple days of class, discuss it with us *as soon as possible*.
- **Live sessions: OPTIONAL BUT HIGHLY ENCOURAGED**
All live, scheduled events for the course are required.
- **Office hours: OPTION OPTIONAL BUT HIGHLY ENCOURAGED**
Office hours are optional, but will be held multiple times per week via Zoom or by appointment.
- **Participating in homework activities: AT LEAST ONCE PER WEEK**
As part of your participation, each week you can expect to post homework or case study solutions.

COURSE MATERIALS AND TECHNOLOGIES

Textbooks

RECOMMENDED/OPTIONAL

All textbooks in this list are available online – either freely available or through the OSU library's Safari website:

1. Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey Ullman, 2014 (Free material available at <http://www.mmds.org/>)
2. Introduction to data mining. Tan, Pang-Ning, Michael Steinbach, Anuj Karpatne, and Vipin Kumar. 2019 (An electronic version is available for online reading from the OSU Safari website).
3. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, by Aurélien Géron, O'Reilly Media, 2017. (An electronic version is available for online reading from the OSU Safari website).
4. Learning Spark: Lightning-Fast Big Data Analysis, Matei Zaharia, Holden Karau, Andy Konwinski, Patrick Wendell, O'Reilly, 2015. (An electronic version is available for online reading from the OSU Safari website).
5. An Introduction to Statistical Learning with Applications in R, James, Witten, Hastie, Tibshirani, Springer, 2013. (Freely downloadable PDF available at <http://www.bcf.usc.edu/~gareth/ISL>).

Course technology

For help with your password, university email, Carmen, or any other technology issues, questions, or requests, contact the OSU IT Service Desk. Standard support hours are available at <https://ocio.osu.edu/help/hours>, and support for urgent issues is available 24/7.

- **Self-Service and Chat support:** <http://ocio.osu.edu/selfservice>
- **Phone:** 614-688-HELP (4357)
- **Email:** 8help@osu.edu
- **TDD:** 614-688-8743

BASELINE TECHNICAL SKILLS FOR ONLINE COURSES

- Basic computer and web-browsing skills
- Navigating Carmen: for questions about specific functionality, see the [Canvas Student Guide](#).

REQUIRED TECHNOLOGY SKILLS SPECIFIC TO THIS COURSE

- Carmen Zoom
- Microsoft Teams
- Basic python programming, ability to use Google Collab or Jupyter Labs framework.

REQUIRED EQUIPMENT

- Computer: current Linux, Mac (OS X), or PC (Windows 7+) system with high-speed internet connection or an iPad.
- Other: a mobile device (smartphone or tablet) or landline to use for BuckeyePass authentication

RECOMMENDED (BUT OPTIONAL) EQUIPMENT

- Webcam: built-in or external webcam, fully installed and tested
- Microphone: built-in laptop or tablet mic or external microphone

REQUIRED SOFTWARE

- Anaconda or access to Google Collab
- R and R Studio

CARMEN ACCESS

You will need to use [BuckeyePass](#) multi-factor authentication to access your courses in Carmen. To ensure that you are able to connect to Carmen at all times, it is recommended that you take the following steps:

- Register multiple devices in case something happens to your primary device. Visit the [BuckeyePass - Adding a Device](#) help article for step-by-step instructions.
- Request passcodes to keep as a backup authentication option. When you see the Duo login screen on your computer, click “Enter a Passcode” and then click the “Text me new codes” button that appears. This will text you ten passcodes good for 365 days that can each be used once.
- Download the [Duo Mobile application](#) to all of your registered devices for the ability to generate one-time codes in the event that you lose cell, data, or Wi-Fi service.

If none of these options will meet the needs of your situation, you can contact the IT Service Desk at 614-688-4357 (HELP) and the IT support staff will work out a solution with you.

GRADING AND FACULTY RESPONSE

How your grade is calculated

ASSIGNMENT CATEGORY	POINTS
HOMEWORK	75%
FINAL PROJECT	25%
Total	100%

Late assignments

As this is a graduate level course, we expect assignments to be submitted on the date they are due before the time indicated. If you find that you are going to be submitting an assignment late it is your responsibility to communicate with your instructor to determine if an extension can be granted.

Grading scale

93–100: A
 90–92.9: A-
 87–89.9: B+

83–86.9: B
80–82.9: B-
77–79.9: C+
73–76.9: C
70 –72.9: C-
67 –69.9: D+
60 –66.9: D
Below 60: E

Faculty feedback and response time

I am providing the following list to give you an idea of my intended availability throughout the course. (Remember that you can call **614-688-HELP** at any time if you have a technical problem.)

- **Grading and feedback:** For large weekly assignments, you can generally expect feedback within **7 days**.
- **Email:** I will reply to emails within **24 hours on days when class is in session at the university**.
- **Discussion board:** I will check and reply to messages in the discussion boards every **24 hours on school days**.

OTHER COURSE POLICIES

Discussion and communication guidelines

The following are my expectations for how we should communicate as a class. Above all, please remember to be respectful and thoughtful.

- **Writing style:** While there is no need to participate in class discussions as if you were writing a research paper, you should remember to write using good grammar, spelling, and punctuation. A more conversational tone is fine for non-academic topics.
- **Tone and civility:** Let's maintain a supportive learning community where everyone feels safe and where people can disagree amicably. Remember that sarcasm doesn't always come across online.
- **Citing your sources:** When we have academic discussions, please cite your sources to back up what you say. (For the textbook or other course materials, list at least the title and page numbers. For online sources, include a link.)

- **Backing up your work:** Consider composing your academic posts in a word processor, where you can save your work, and then copying into the Carmen discussion.

Academic integrity policy

OHIO STATE'S ACADEMIC INTEGRITY POLICY

Academic integrity is essential to maintaining an environment that fosters excellence in teaching, research, and other educational and scholarly activities. Thus, The Ohio State University and the Committee on Academic Misconduct (COAM) expect that all students have read and understand the University's [*Code of Student Conduct*](#), and that all students will complete all academic and scholarly assignments with fairness and honesty. Students must recognize that failure to follow the rules and guidelines established in the University's *Code of Student Conduct* and this syllabus may constitute "Academic Misconduct."

The Ohio State University's *Code of Student Conduct* (Section 3335-23-04) defines academic misconduct as: "Any activity that tends to compromise the academic integrity of the University, or subvert the educational process." Examples of academic misconduct include (but are not limited to) plagiarism, collusion (unauthorized collaboration), copying the work of another student, and possession of unauthorized materials during an examination. Ignorance of the University's *Code of Student Conduct* is never considered an excuse for academic misconduct, so I recommend that you review the *Code of Student Conduct* and, specifically, the sections dealing with academic misconduct.

If we suspect that a student has committed academic misconduct in this course, we are obligated by University Rules to report suspicions to the Committee on Academic Misconduct. If COAM determines that you have violated the University's *Code of Student Conduct* (i.e., committed academic misconduct), the sanctions for the misconduct could include a failing grade in this course and suspension or dismissal from the University.

If you have any questions about the above policy or what constitutes academic misconduct in this course, please contact me.

Other sources of information on academic misconduct (integrity) to which you can refer include:

- The Committee on Academic Misconduct web pages ([COAM Home](#))
- *Ten Suggestions for Preserving Academic Integrity* ([Ten Suggestions](#))
- *Eight Cardinal Rules of Academic Integrity* (www.northwestern.edu/uacc/8cards.htm)

Artificial Intelligence Policy

Artificial intelligence tools including, but not limited to ChatGPT or Microsoft Copilot can be useful tools for summarizing ideas and brainstorming, but it is not permitted to copy and submit text that was predominantly generated by AI or use code directly generated by AI. All work submitted must ultimately be your own.

Copyright disclaimer

The materials used in connection with this course may be subject to copyright protection and are only for the use of students officially enrolled in the course for the educational purposes associated with the course. Copyright law must be considered before copying, retaining, or disseminating materials outside of the course.

Statement on Title IX

Title IX makes it clear that violence and harassment based on sex and gender are Civil Rights offenses subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories (e.g., race). If you or someone you know has been sexually harassed or assaulted, you may find the appropriate resources at <http://titleix.osu.edu> or by contacting the Ohio State Title IX Coordinator, Kellie Brennan, at titleix@osu.edu

Your mental health

A recent American College Health Survey found stress, sleep problems, anxiety, depression, interpersonal concerns, death of a significant other, and alcohol use among the top ten health impediments to academic performance. Students experiencing personal problems or situational crises during the quarter are encouraged to contact Ohio State University Counseling and Consultation Service (614-292-5766; www.ccs.osu.edu) for assistance, support and advocacy. This service is free and confidential.

ACCESSIBILITY ACCOMMODATIONS FOR STUDENTS WITH DISABILITIES

Requesting accommodations

If you would like to request academic accommodations based on the impact of a disability qualified under the Americans with Disabilities Act and Section 504 of the Rehabilitation Act of 1973, contact your instructor privately as soon as possible to discuss your specific needs. Discussions are confidential.

In addition to contacting the instructor, please contact the Student Life Disability Services at [614-292-3307](tel:614-292-3307) or ods@osu.edu to register for services and/or to coordinate any accommodations you might need in your courses at The Ohio State University.

Go to <http://ods.osu.edu> for more information.

Accessibility of course technology

This online course requires use of Carmen (Ohio State's learning management system) and other online communication and multimedia tools. If you need additional services to use these technologies, please request accommodations with your instructor.

- [Carmen \(Canvas\) accessibility](#)
- Streaming audio and video
- Synchronous course tools

COURSE SCHEDULE

This schedule is **tentative** and may change as the semester progresses. The actual schedule will be posted and maintained in Carmen.

Lecture	Topic
Week 1	Hierarchical and agglomerative clustering methods
Week 2	K-Means and BIRCH
Week 3	Classification and Regression Trees
Week 4	Predictions and Predictive Errors
Week 5	Random Forest

Week 6	K-Nearest Neighbors and Naive Bayes
Week 7	Statistical learning frameworks; Unsupervised vs. supervised learning; Popular supervised methods including penalized regression (Ridge Regression, LASSO, and Elastic Net)
Week 8	Support vector machines and kernel methods
Week 9	Definitions of error/loss, learning, and generalization; Bias vs. variance trade-offs; Model evaluation (with test data, cross validation, bootstrap); Averaging prediction across models; Addressing training and test data mismatch.
Week 10-12	Neural networks, Deep learning, Transfer learning. Deployment on multitude of data and context-driven evaluations.