Section Syllabus

CSE 5243

Introduction to Data Mining  
Autumn 2022

**Health and safety requirements:** All students, faculty and staff are required to comply with and stay up to date on all university safety and health guidance (<https://safeandhealthy.osu.edu>), which includes following university mask policies and maintaining a safe physical distance at all times. Non-compliance will be warned first and disciplinary actions will be taken for repeated offenses. (Updated: Aug. 14, 2020)

# Course overview

## Instructor

Instructor: Gregory Ryslik

Email address: [ryslik.1@osu.edu](mailto:ryslik.1@osu.edu)

Phone number: N/A – use email for default communication medium

Office hours: TBD – to be posted on Carmen/Github when semester begins

## Course description

See course syllabus (<http://coe-portal.cse.ohio-state.edu/pdf-exports/CSE/CSE-5243.pdf>).

## Course learning outcomes

See course syllabus (<http://coe-portal.cse.ohio-state.edu/pdf-exports/CSE/CSE-5243.pdf>).

# How this course works

**Mode of delivery:** This course is a hybrid course. All lectures will be recorded for those that want to attend virtually. I will let the course know ahead of time if the course recording will be virtual or in person at least one week in advance based upon my travel schedule.

**Pace of activities:** This course is divided into **modules** that are released incrementally throughout the semester. Students are expected to attend lectures and keep pace with assignment deadlines.

**Credit hours and work expectations:** This is a **3-credit-hour course**. According to [Ohio State policy](http://go.osu.edu/credithours), 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.

**Homework assignments:** Students are expected to complete five homework assignments during the semester. These homework assignments may vary slightly from semester to semester, but typically cover data mining algorithm programming in Python, and general analysis of data to solve problems. The assignments usually cover data preprocessing, classification, clustering, association mining, and optional topics such as text mining and recommendation systems. Students are expected to develop and test Python code and/or use Excel or related tools to perform the analysis, and to submit a report of their findings.

**Attendance and participation requirements:** Consistent engagement is expected. If any problems arise relative to attendance, please contact the instructor as soon as possible. Communication is important. You are encouraged to participate “in class”, ask questions, work on in-class problems in small groups, and share your experiences relative to the subjects and discussion that day.

The lectures will be interactive and will be helpful for you to be successful in mastering the course learning outcomes. Attendance and active participation often impact your performance in a meaningful way, so it will be beneficial for you to attend this course synchronously. The lecture slides will be posted on CarmenCanvas, so if you do miss a lecture, you are expected to view the missed material before the next lecture. The following is a summary of everyone's expected participation:

* **Office hours**:  **OPTIONAL**. Office hours will be via CarmenZoom. Students will enter the CarmenZoom waiting room and will be served one at a time or in groups, depending on the circumstances.

# Course materials and technologies

## Course materials (slide decks, readings, etc.) will be provided by the instructor.

## Textbooks

### Required

* None

### Recommended/optional

* (Primary) Data Mining: Concepts and Techniques, 3rd edition, Morgan Kaufmann, Jiawei Han, Micheline Kamber, and Jian Pei.
* (Primary) Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, and Vipin Kumar
* (Supplementary) Data Mining Analysis and Concepts (Online version available), Mohammed J. Zaki and Wagner Meira, Jr.
* (Supplementary) Mining of Massive Datasets (Online version available), Jure Leskovec, Anand Rajaraman and Jeffrey Ullman
* (Supplementary) Machine Learning, Tom Mitchell  
  (Supplementary) Pattern Recognition and Machine Learning, Christopher M. Bishop

## Other fees or requirements

* None

## 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](mailto:8help@osu.edu)
* **TDD:** 614-688-8743

### Baseline technical skills for courses

* Basic computer and web-browsing skills
* Navigating Carmen: for questions about specific functionality, see the [Canvas Student Guide](https://community.canvaslms.com/docs/DOC-10701).

### Required Technology skills specific to this course

* CarmenConnect (CarmenZoom) text, audio, and video chat
* Recording a slide presentation with audio narration
* Recording, editing, and uploading video

### Required equipment

* Computer: current Linux, Mac (macOS 10.13+), or PC (Windows 8+) system with high-speed internet connection.
* Authentication device: a mobile device (smartphone or tablet), or landline, or security key (e.g., YubiKey, Feitian) to use for BuckeyePass authentication.
* Scanner: a camera, smartphone, tablet, or document scanner for scanning and uploading hand-written documents such as homeworks and exams.
* Webcam: built-in or external webcam, fully installed and tested.
* Microphone: built-in laptop or tablet mic or external microphone.

### Required software

* [Microsoft Office 365](https://ocio.osu.edu/blog/community/2015/08/18/free-microsoft-office-for-ohio-state-students): All Ohio State students are eligible for free Microsoft Office 365 ProPlus through Microsoft’s Student Advantage program. Full instructions for downloading and installation can be found at [go.osu.edu/office365help](http://go.osu.edu/office365help). In particular, some assignments and in-class exercises may use Excel, so it is important that students have access to Excel.
* Python: Assignments will be done in Python. The preferred environment is Anaconda (<http://www.anaconda.org>). In particular, the strongly preferred modality for assignment submissions is Jupyter Notebook, which is included in the Anaconda environment.
* Discussion Board: We will make use of the class-wide discussion board on Carmen to discuss questions/issues/etc.

### carmen access

You will need to use [BuckeyePass](https://buckeyepass.osu.edu/) 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](https://osuitsm.service-now.com/selfservice/kb_view.do?sysparm_article=kb05025) 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](https://osuitsm.service-now.com/selfservice/kb_view.do?sysparm_article=kb05026) 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 x 5 | 60 |
| Participation / Attendance | 5 |
| Midterm Exam | 15 |
| Final Exam | 20 |
| Total | **100** |

*See course schedule below for due dates.*

## Late assignments

Late submissions and missed exams **will not be accepted**, as a default. If extenuating circumstances arise, talk with the instructor as soon as possible. Please refer to Carmen for due dates.

## 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 assignments and exams, you can generally expect feedback from the course TA, or me, within **14 days (typically sooner)**.
* **Email:** I will reply to emails usually within **48 hours on days when class is in session at the university (typically sooner)**.
* **Discussion board:** The TA or I will check and reply to messages in the discussion board every **48 hours on school days (typically sooner)**. Please make use of the discussion board – it is a great resource for getting answers to questions on the assignments, etc. Please be proactive and ask your questions as early as possible – do not wait until the assignment is almost due before asking for clarifications.

# 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

### Policies for this course

* **Quizzes and exams**: You must complete the midterm and final exams yourself, without any external help or communication. Periodic quizzes may be included as self-checks without points attached.
* **Written assignments**: Your written assignments, including discussion posts, should be your own original work. In formal assignments, you should follow MLA or APA style to cite the ideas and words of your research sources. You are encouraged to ask a trusted person to proofread your assignments before you turn them in—but no one else should revise or rewrite your work.
* **Reusing past work**: In general, you are prohibited in university courses from turning in work from a past class to your current class, even if you modify it. If you want to build on past research or revisit a topic you've explored in previous courses, please discuss the situation with me.
* **Falsifying research or results**: All research you will conduct in this course is intended to be a learning experience; you should never feel tempted to make your results or your library research look more successful than it was.
* **Collaboration and informal peer-review**: The course includes many opportunities for formal collaboration with your classmates. While study groups and peer-review of major written projects is encouraged, remember that comparing answers on a quiz or assignment is not permitted. If you're unsure about a particular situation, please ask ahead of time.
* **Group projects**: This course does not include graded group projects. However, we may consider the option of allowing multi-person teams to collaborate on some homework assignments. If we do this (at the instructor’s discretion), it will be optional for the students.

### 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*](https://studentconduct.osu.edu/), 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 I suspect that a student has committed academic misconduct in this course, I am obligated by University Rules to report my 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](http://oaa.osu.edu/coam.html))
* Ten Suggestions for Preserving Academic Integrity ([*Ten Suggestions*](http://oaa.osu.edu/coamtensuggestions.html))
* Eight Cardinal Rules of Academic Integrity ([www.northwestern.edu/uacc/8cards.htm](http://www.northwestern.edu/uacc/8cards.html))

## 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](mailto: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 or [ods@osu.edu](mailto: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](http://ods.osu.edu/) for more information.

## Accessibility of course technology

This 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](https://community.canvaslms.com/docs/DOC-2061)
* Streaming audio and video
* Synchronous course tools

# Course schedule

The course schedule below is the approximate sequencing of the curriculum. Minor changes may be made to accommodate guest speakers, respond to common student needs, etc.

| Week | Topics, Readings, Assignments, Deadlines |
| --- | --- |
| 1 | Course Introduction |
| 2 | Review of Basic Probability and Statistical Concepts |
| 3 | Data and Data Preprocessing |
| 4 | Data and Data Preprocessing |
| 5 | Classification |
| 6 | Classification |
| 7 | Classification |
| 8 | Classification |
| 9 | Midterm Exam, Clustering |
| 10 | Clustering |
| 11 | Clustering, Frequent Patterns and Associations |
| 12 | Frequent Patterns and Associations |
| 13 | Frequent Patterns and Associations |
| 14 | Anomaly Detection |
| 15 | Misc Topics (e.g.): Graphs / PageRank, Text Mining, Recommendation Systems |
| 16 | Final Exam |