



LOYOLA UNIVERSITY CHICAGO

COMP 406: Data Mining Fall 2024

Wed, 5:30 pm – 8 pm CT
Online via [Zoom](#)

Please note that this syllabus may be subject to changes within the first few weeks of classes, mainly to enable minor corrections. Please check Sakai for any updates. Thank you for your understanding.

Course Information

Course Number	COMP 406
Course Name	Data Mining
Course Description	This course covers theory and practice of the analysis (mining) of extremely large datasets. With data growing at exponential rates knowledge gathering and exploration techniques are essential for gaining useful intelligence.
Modality	Online (synchronous)
Course Duration	Aug 28, 2024 – Dec 14, 2024
Lecture Time	Wednesdays, 5:30 PM – 8 PM CT
Zoom Link	https://luc.zoom.us/j/82087461309?pwd=4Th9pPMWa3bDoco9TJhFtA7NOAm67w.1

Prerequisites

(COMP 231 or COMP 251 or COMP 271) and (STAT 103 or STAT 203 or ISSCM 241 or PSYC 304 or instructor permission)

Helpful Links

- [Lectures Zoom Link](#)
- [Office Hours](#)
- [Gradescope](#) (code: 2BBEZ3)
- [Piazza](#)
- [Late Submission Form](#)

Instructor Information

Instructor

Dr. Satyaki Sikdar

Campus Office

211 Doyle Center

Email

ssikdar@luc.edu

Office Hours

Online and in-person by appointment via [Calendly](#)

Contact Preference

Piazza => Email

Course Components

Learning Outcomes

Upon successful completion of this course, you will be able to:

- Understand various kinds of structured and unstructured data, including relational, transactional, graph, and spatial data.
- Realize the various steps involved in the Extract, Transform, and Load (ETL) framework using Pandas.
- Visualize datasets and learn fundamental design principles and best practices using Matplotlib and Seaborn.
- Apply preprocessing and cleaning procedures on messy real-world data.
- Extract meaningful patterns from transactional data using frequent pattern mining.
- Distinguish between supervised and unsupervised learning techniques.
- Use fundamental supervised learning models like decision trees, k-nearest neighbors, and neural networks on synthetic and real-world datasets.
- Find patterns in data using unsupervised techniques like k-means clustering, hierarchical, and density-based methods.
- Demonstrate a thorough grasp of the concepts by completing more complex tasks, including but not limited to working with multiple data sources, formulating research questions, and gathering empirical evidence to support primary hypotheses.
- Generate publication-quality figures and develop formal research writing skills.

Deliverables

Problem Sets

Problem sets will be released on **Thursdays at 9 am** and due the following **Wednesday at 3 pm**. The lowest-scoring problem set (excluding PS 0) will be dropped when calculating the final grade. Over the semester, 7-8 problem sets will be released.

Problem sets may have additional challenging questions for graduate students.

Quizzes

There will be 3 timed online quizzes administered on **Sakai**. There may be some overlap in content across quizzes. Each quiz will be worth 50 points and contain a mix of multiple-choice and short-answer questions. They will be closed book and closed internet, but you can use your class notes. You are expected to solve the problems on your own. Collaborating with other students or using other resources will be considered an Honor Code violation.

Tentatively, Quiz 1 will be held between Sep 19 and Sep 24, Quiz 2 between Oct 10 and Oct 15; and Quiz 3 between Nov 21 and 25. The actual dates and topics covered will be posted on Sakai.

Quizzes may have additional challenging questions for graduate students.

Final Exam

There will not be a final exam. The final course project deliverables will be treated as a final.

Participation

After each lecture, you will be asked to write a brief reflection statement (4-5 sentences) covering the topics discussed. These will be administered on **Sakai** and will be due **24 hours** after the end of the lecture. These will be graded for completion only, and the two lowest-scoring assignments will be dropped while calculating the final grade.

Course Project

There will be a semester-long course project. You will work in teams of 3-4 with your classmates. There is only one restriction while forming teams: everyone on your team must be in the same cohort, i.e., COMP 306 takers cannot team up with those taking COMP 406.

There will be recurring checkpoints throughout the semester to track each group's progress. A final project report and a 10 min recorded presentation are due at the end of the semester during the week of final exams. More details will be updated on the course Sakai page.

Graduate students are expected to produce more thoughtful and more challenging course projects.

Tools

Sakai

Sakai is the online course management system we will use to run the class. All course content, including the syllabus, lecture videos, assignments, solutions, and other course materials, will be posted here. It will also host your gradebook, which will be updated at least once a month. Please check the Sakai page at least once a week.

Gradescope

We will use Gradescope to administer the weekly problem sets. You can access Gradescope via the Assignments page on Sakai. For more information about Gradescope, please visit this [website](#).

Zoom

Weekly lectures will be hosted on Zoom. All lecture recordings will be posted on Sakai.

Google Colab

The problem sets will be administered on the [Google Colab](#) platform.

Piazza

We will use [Piazza](#) to host all out-of-class communications. This is also the preferred way to contact the course staff via DMs.

If you have questions about anything related to the course, please post them on the Piazza class feed. Please do not post anything resembling a solution to a homework problem before it is due. If in doubt, you should make your post visible to only instructors and TAs. You are also welcome to remain Anonymous in your posts or responses. We always welcome any feedback on what we could be doing better.

Piazza signup link: <https://piazza.com/luc/fall2024/comp306406>

Tentative Schedule

This schedule may be altered to suit the needs of the class. Please check the Sakai page for the latest updates.

Week #	Date	Topic	Deliverables
1	Wed, Aug 28	Course Introduction, ETL, Python 101	
2	Wed, Sep 4	Data Preprocessing - I	PS 0
3	Wed, Sep 11	Data Preprocessing - II	PS 1
4	Wed, Sep 18	Correlations and Distance Measures	PS 2. Quiz 1. Project teams.
5	Wed, Sep 25	Frequent Pattern Mining	
6	Wed, Oct 2	Supervised Learning - I	PS 3. Project proposals.
7	Wed, Oct 9	Supervised Learning - II	PS 4. Quiz 2.
8	Wed, Oct 16	Supervised Learning - III	PS 5
9	Wed, Oct 23	Imbalanced classification, AUC, AUROC, AUPR	Project check-ins.
10	Wed, Oct 30	Unsupervised Learning - I	PS 6
11	Wed, Nov 6	Unsupervised Learning - II	PS 7
12	Wed, Nov 13	Unsupervised Learning - III	PS 8
13	Wed, Nov 20	Course review	Quiz 3
14	Wed, Nov 27	Thanksgiving break	
15	Wed, Dec 4	Advanced topics	Project presentations
16	Week of Dec 9	Finals week	Project reports

Requirements and Expectations

Course Textbooks

Data Mining: Concepts and Techniques (required)

Jiawei Han, Micheline Kamber, and Jian Pei

3rd Edition, published in 2012.

ISBN: 1-283-17117-1

The book is freely available online on Ebook Central via the [Loyola Libraries](#). You can also look for physical copies on Amazon, the international version typically retails for \$30-40.

Introduction to Data Mining (optional)

Pan-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar

2nd Edition (international / global), published in 2019.

ISBN: 978-9354491047

The international version typically retails for \$20-30 on Amazon.

What I expect of you

- That you are fully present when participating in the class.
- That you abide by the Loyola University Honor Code.
- That you are kind to yourself, your fellow students, and the course staff.
- That, if something happens to interfere with your ability to meet your course obligations, you tell me as soon as possible.
- You communicate professionally in all class-related communications - email, text, instant messages, etc., adhering to standard Internet rules of netiquette.
- That you do the assigned activities thoughtfully and on time.
- That you have fun learning about Data Mining.

What you can expect of me

- That I give you coursework that is well organized and reasonable to schedule your study time appropriately.
- That I give you challenging work that allows you to explore the topic in a way that interests you.
- That I give you regular graded activities so that you always know how you are doing in the class.
- That I give you advanced guidance on how things will be graded, so you know my expectations for an assignment.
- That I grade the work in a timely manner (within 10 days of the due date) and that I grade fairly.
- That I give you all the support and training you need to succeed in the classroom.
- That I have fun teaching the class so that it will be fun for you to learn.

Evaluation and Grading

Grade Distribution

Component	Weight
Problem Sets	30%
Quizzes	3 x 10% = 30%
Participation	10%
Course Project	30%

Scale

A	A-	B+	B	B-	C+	C	C-	D+	D	F
94+	90+	87+	84+	80+	77+	74+	70+	67+	60+	<60

Academic Integrity

Loyola University Chicago takes seriously the issues of plagiarism and academic integrity. Below is an excerpt, quoted directly, of the university's statement on integrity.

"The faculty and administration of Loyola University Chicago wish to make it clear that the following acts are regarded as serious violations of personal honesty and the academic ideal that binds the university into a learning community:

Submitting as one's own:

1. Material copied from a published source: print, internet, CD-ROM, audio, video, etc.
2. Another person's unpublished work or examination material.
3. Allowing another or paying another to write or research a paper for one's own benefit.
4. Purchasing, acquiring, and using for course credit a pre-written paper.

The critical issue is to give proper recognition to other sources. To do so is both an act of personal, professional courtesy and of intellectual honesty."

Read through [Loyola's full statement on Academic Integrity](#) here.

These policies fully apply to this course. The penalty for task-wise academic

misconduct is losing all the task's points. Multiple events of misconduct will be reported through the proper channels.

Using generative AI tools

You are not allowed to use any AI-assisted technology (like ChatGPT) during the course, unless explicitly authorized by the instructor.

Grade Change and Regrade Requests

Any requests for a grade change on an assignment or a quiz must be made within **seven** days when the grades are published. Any clerical errors while grading will be happily corrected.

All change requests for problem sets must be made on Gradescope. Grade changes for quizzes must be made as a **private post** on Piazza, visible only to the student and the instructor.

Course Policies

General

- Please be kind to yourself, your fellow students, and the course staff.
- One important aspect of a Jesuit education is learning to respect the rights and opinions of others. Please respect others by (1) allowing all classmates the right to voice their opinions without fear of ridicule and (2) not using profanity or making objectionable (gendered, racial or ethnic) comments, especially comments directed at a classmate.
- Class rosters are provided to each instructor with the student's legal name. I will gladly honor your request to address you by an alternate name and/or gender pronoun.
- Life happens, and problems may come up which may interfere with your life and work. If you feel the need to sort something out related to this class, please contact Prof. Sikdar. If you have a problem with Prof. Sikdar, please contact the department chair, Prof. George Thiruvathukal.
- I am committed to the principle of universal learning. This means that our classroom, virtual spaces, practices, and interactions are as inclusive as possible. Mutual respect, civility, and the ability to listen and observe others are crucial to universal learning. Active, thoughtful, and respectful participation in all aspects of the course will make our time together as productive and engaging as possible.

Attendance

The class will be conducted synchronously online over Zoom. You are expected to attend all weekly lectures.

At Loyola, faculty members will make efforts to accommodate students if the observance of a major religious holiday interferes with a student's academic work. If you are unable to attend a class, participate in a test or quiz, be present for a presentation, or complete an assignment on a specific day because of the observance of a religious holiday, you will be excused and provided the opportunity to make up the work. You will continue to be responsible for all assigned work and should notify the professor in advance through Loyola email of the religious observance(s) that conflict with your classes.

Student Participation

- You are expected to check the course Sakai page weekly for any updates.
- Please strive to be your best self in this online context: be respectful and patient with the instructor and your classmates. The LUC Community Standards also apply online. Students must actively engage in course activities and discussions during synchronous sessions. The following list details my expectations for you in our synchronous sessions:
 - Login: Join Zoom using your LUC Zoom account. You can update your name or add your pronouns through your Zoom profile.
 - Device use: You must join the weekly lectures on Zoom from a computer and **not** a cell phone or tablet.
 - Camera use: Turn your camera on during course discussions and activities. Video use in synchronous meetings can foster a sense of community which leads to more engaging discussions. Additionally, turning your video off may make it more difficult for your peers to understand you during discussions, especially for peers with hearing impairments. Turn your video off when you need a quick break, your background may be distracting, or you encounter connection issues.
 - Background: Consider using a virtual background for your privacy and to limit distractions during synchronous sessions.

- Audio: Unmute your microphone to answer questions and engage in class discussions. Mute your microphone when you are not speaking to reduce background noise. Consider using a headset or headphones with a microphone to improve audio quality. When possible, try to connect in a quiet, distraction-free place.
- Chat: Use the chat window for questions and comments that are relevant to our class. Comments that are off topic make it difficult for me to address students' questions about the course. Disrespect or hate speech will not be tolerated.
- Session Recordings: Do not share class session recordings with anyone who is not currently enrolled in this course section. Unauthorized student recording of classroom activities is prohibited. Students who engage in unauthorized recording, unauthorized use of a recording, or unauthorized distribution of instructional materials will be subject to the processes and procedures that uphold Community Standards. To request the use of assistive technology for an accommodation should direct such requests to the Student Accessibility Center.

Late work and make up

Life often gets in the way of study. To accommodate that, you have 2 Late Passes with no questions asked. To use a Late Pass, you must:

- Notify Prof. Sikdar **before** the assignment is due. Any time before the deadline is fine. The notification must be made via the [late pass submission form](#). Your request will be approved as long as you have late passes remaining. Thus, you do not need to wait for a response.
- Turn in the assignment within 3 calendar days of the original due date. Please include a note that it is being submitted late with a Late Pass.
- Assignments turned in with a late pass after the 3-day grace period will be penalized at 10% / day up for a maximum of 4 days. No late work will be accepted beyond that.
- You may only use one Late Pass per assignment.
- The Late Passes **will not** be applied automatically or retroactively to a past assignment.
- If no late pass is used, any late work will be accepted for a **maximum of 7 days** with a 10% penalty/day. No late work will be accepted after that.

Please note: the late passes are here to make the class easier for you. There is no benefit to not using the late pass, so please use them. That is why they are available.

Class Recording Notification

This course will be recorded using Zoom. This system allows us to automatically record and distribute lectures to you in a secure environment. You can watch these recordings on your computer, tablet, or smartphone. In Sakai, look for the links to the lecture videos on the course schedule. Because we will be recording the lectures, your questions and comments may be recorded. Except for faculty and staff who require access, no content will be shared with individuals outside your course without your explicit permission. Loyola University Chicago and your instructor jointly copyright these recordings. Posting them to other websites (including YouTube, Facebook, Snapchat, etc.) or elsewhere without express written permission may result in disciplinary action and possibly civil prosecution.

Privacy Statement

Assuring privacy among faculty and students engaged in online and face-to-face instructional activities helps promote open and robust conversations and mitigates concerns that comments made within the context of the class will be shared beyond the classroom. As such, recordings of instructional activities occurring in online or face-to-face classes may be used solely for internal class purposes by the faculty member and students registered for the course, and only during the period in which the course is offered. Students will be informed of such recordings by a statement in the syllabus for the course in which they will be recorded. Instructors who wish to make subsequent use of recordings that include student activity may do so only with informed written consent of the students involved or if all student activity is removed from the recording. Recordings, including student activity initiated by the instructor, may be retained by the instructor only for individual use.

Intellectual Property

All lectures, notes, PowerPoints, and other instructional materials in this course are the intellectual property of the professor. As a result, they may not be distributed or shared in any manner, either on paper or virtually, without my written permission. Lectures may not be recorded without my written consent; when consent is given, those recordings may be used for review only and may not be distributed. Recognizing that your work, too, is your intellectual property, I will not share or distribute your work in any form without your written permission.

Statement of Intent

By remaining in this course, students are agreeing to accept this syllabus as a contract and to abide by the guidelines outlined in the document. Students will be consulted should there be a necessary change to the syllabus.

Notice of Reporting Obligations for Responsible Campus Partners

As an instructor, I am a Responsible Campus Partner ("RCP") under Loyola's [Comprehensive Policy and Procedures for Addressing Discrimination, Sexual Misconduct, and Retaliation](#) (available at www.luc.edu/equity). While my goal is for you to be able to engage fully and authentically with our course material through class discussions and written work, I also want to be transparent that as a RCP I am required to report certain disclosures of sexual misconduct (including sexual assault, sexual harassment, intimate partner and/or domestic violence, and/or stalking) to the [Office for Equity & Compliance](#) ("OEC"). As the University's [Title IX](#) office, the OEC coordinates the University's response to reports and complaints of sexual misconduct (as well as discrimination of any kind) to ensure students' rights are protected.

As an instructor, I also have an obligation under Illinois law to report disclosures of or suspected instances of child abuse or neglect (<https://www.luc.edu/hr/legal-notices/mandatedreportingofchildabuseandneglect/>).

The University maintains such reporting requirements to ensure that any student who experiences sexual/gender-based violence receives accurate information about available resources and support. Such reports will not generate a report to law enforcement (no student will ever be forced to file a report with the police). Additionally, the University's resources and support are available to all students even if a student chooses that they do not want any other action taken. If you have any questions about this policy, you are encouraged to contact the OEC at

equity@luc.edu or 773-508-7766.

If you ever wish to speak with a confidential resource regarding gender-based violence, I encourage you to call [The Line](#) at 773-494-3810. The Line is staffed by confidential advocates from 8:30am-5pm M-F and 24 hours on the weekend when school is in session. Advocates can provide support, talk through your options (medical, legal, LUC reporting, safety planning, etc.), and connect you with resources as needed – *without* generating a report or record with the OEC. More information about The Line can be found at luc.edu/wellness.

Student Accommodations

Loyola University provides reasonable accommodations for students with disabilities. Any student requesting accommodations related to a disability or other condition is required to register with Student Accessibility Center (SAC), located in Sullivan Center, Suite 117. Professors receive the accommodation notification from SAC via Accommodate. Students are encouraged to meet with their professor individually in order to discuss their accommodations. All information will remain confidential. Please note that in this class, software may be used to record class lectures in order to provide equal access to students with disabilities. Students approved for this accommodation use recordings for their personal study only and recordings may not be shared with other people or used in any way against the faculty member, other lecturers, or students whose classroom comments are recorded as part of the class activity. Recordings are deleted at the end of the semester. For more information about registering with SAC or questions about accommodations, please contact SAC at 773-508-3700 or SAC@luc.edu.

Mental Health

In a given year, 5-20% of US residents will contract influenza (this rate varies greatly depending on the vaccine's effectiveness and virulence of the seasonal strain). In contrast, around 7% of US residents experience a major depressive episode each year. In other words, depression is as common as the flu, yet one can sometimes be socially stigmatized, and the other is not. From my point of view as your professor, if you are having mental health issues that are interfering with your ability to function (in my course or otherwise), that is an entirely valid medical reason to ask for short term extensions or accommodations, just as you might for a physical illness or injury. So, if your situation is hard or you feel

hopeless, do not hesitate to reach out, and I will try to work with you to minimize the impact mental health difficulties have on your performance in my class.

During the semester, if you find that health problems, life stressors or emotional difficulties are interfering with your academic or personal success, and you are therefore finding it difficult to cope or to complete your academic work, please consider contacting the Wellness Center. Healthcare services, crisis intervention, time-limited individual counseling, and group therapies are free of charge, and strictly confidential, having nothing to do with your educational records.

You can make an appointment [online](#). You may also call 773-508-2530 for counseling appointments or 773-508-8883 to speak with a nurse about medical concerns. More information is available at the [Wellness Center](#). If your medical or mental health condition requires ongoing academic accommodations, please register with [Student Accessibility Services](#) and provide me with a copy of your accommodation letter.
