Syllabus for Knowledge Discovery and Data Mining

CIS 635 (Section 03)   Winter 2025

In this course, we'll delve into the world of data mining, uncovering valuable insights from vast datasets. We will explore techniques to identify meaningful patterns, correlations, and trends, and apply them to both real-world and synthetic data. Topics covered include data preprocessing, dimensionality reduction, data visualization, predictive modeling, model evaluation, clustering, and association analysis techniques.

Contact Information:

Instructor:	Dr.Yong Zhuang
E-mail:	yong.zhuang@gvsu.edu
Office:	MAK D-2-234
Office Hours:	**Monday** 3:00 pm - 5:00 pm, remote (Zoom)
	Or by appointment. Email me to schedule a meeting (over Zoom or in person).
Course Page:	https://gvsu-cis635.github.io & Blackboard
Zoom:	Meeting ID: 396 668 6420, Password: 587684
Section 03:	**Midterm:** The week starting Monday, February 17
	Final exam: The week starting Monday, April 21

Course Objectives:

After completing this course, students should be able to:

▍ Wrangle, analyze and visualize data in **Python**.

▍ Design software pipelines for supervised and unsupervised knowledge discovery.

▍ Identify and build models using common data mining techniques.

▍ Evaluate models for their effectiveness and appropriateness.

▍ Communicate findings using effective visualizations.

Prerequisites:

CIS 500: Fundamentals of Software Practice

Course Materials:

There are no required textbooks for this course. If you would like additional references, some good options include:

✳ *Data Mining Concepts and Techniques (4th Edition)* by Jiawei Han, Jian Pei, and Hanghang Tong. Publication Date: 2023. (freely available online through the GVSU library)

* *Think Python: How to Think Like a Computer Scientist* by Allen B. Downey. (freely available online)
* *Python Data Science Handbook* by Jake VanderPlas. (freely available online)
* *Applied Machine Learning in Python* by Andreas C. Müller. (freely available online)

Slides, recommended readings, online references, and other resources will be made available on Course website & Blackboard.

Course Delivery:

This course will be delivered **in an asynchronous format**

Expectations: I expect the following from you to ensure your success in this course:

a) check Course website & Blackboard regularly for announcements, course material, and assignments

b) adhere to the CIS & GVSU policy of Academic Honesty

Course materials, assignments, grades, and announcements will be posted to Course website & Blackboard (https://lms.gvsu.edu/). It is your responsibility to stay informed.

Note: I use Blackboard's course announcements to communicate timely information about the course (e.g., changes to deadlines, etc). Ensure that your notification settings for Blackboard are set up to allow you to get updates as they are posted.

Grading Proportions:

Course Component	Overall Weight
Quizzes & Homework assignments	30%
Project	30%
Midterm Exam	20%
Final Exam	20%
Total	**100%**

🖊 **Homework assignments** Homework is assigned to reinforce course material, consisting of programming tasks and mathematical or written questions. Unless specified, homework should be completed individually. Sharing code or responses is not permitted. However, discussing high-level concepts and asking questions is encouraged.

🖊 **Project** Students will work in groups on a final project. Each group will choose a dataset and develop a project that applies data mining and knowledge discovery techniques to their selected dataset.

Grade A	Grade B	Grade C	Grades D & F
$A \geq 93\%$	$B+ \geq 87\%$	$C+ \geq 77\%$	$D+ \geq 67\%$
$A- \geq 90\%$	$B \geq 83\%$	$C \geq 73\%$	$D \geq 60\%$
	$B- \geq 80\%$	$C- \geq 70\%$	$F < 60\%$

Course Policies:

Assignments and Due Dates:

☞ **Due dates:** All assignments will be due at 11:59 pm Michigan time on the due date.

☞ **Late policy:** Assignments submitted late will incur a 10% penalty per day, capped at five days (50%). After this period, the assignment will not be accepted.

☞ **Accommodations:** Inform your instructor of any required accommodations to ensure successful learning. Any impediments should be discussed with the instructor.

☞ The instructor reserves the right to modify course policies, calendars, and due dates.

☞ This course adheres to the GVSU policies available at www.gvsu.edu/coursepolicies/.

Academic Honesty:

All students are expected to adhere to the academic honesty standards of Grand Valley State University. In addition, students in this course are expected to adhere to the academic honesty guidelines as set forth by the School of Computing. Details can be found at https://www.gvsu.edu/computing/academic-honesty-30.htm

I believe that you can learn a lot from your peers, both in the class and in the broader community. Therefore, I encourage collaboration with both. However, do not mistake this as a license to cheat. Learning from and with your peers is encouraged, but passing their work off as your own is prohibited. With respect to all individual assignments in this course:

☞ Document all collaborations.

☞ No electronic code transfers between students.

☞ Code you find online must be cited, with an active link to that code. That code should not solve the entirety of an assigned problem/project (i.e., don't have someone else do your project for you).

☞ You are encouraged to engage in conversations in online forums, but do not post solutions or solicit others to complete your work for you.

☞ You are encouraged to talk about problems with each other in non-technical terms (i.e., not code)

☞ Ultimately, you are responsible for all aspects of your submissions. You should be able to explain and defend your submission if the work is entirely your own.

Academic Resources: GVSU also provides opportunities for students to improve their academic skills through resources, such as:

▌ **The writing center:** The Fred Meijer Center for Writing, with locations at the Allendale and Pew/Downtown Grand Rapids campuses, is available to assist you with writing for any of your classes. For more information about these services and locations, please visit their website: http://www.gvsu.edu/wc/

▌**Speech lab:** The Grand Valley Speech Lab is a peer-to-peer communication center that helps students with all elements of oral presentations. For more information about this service, please visit their website: https://www.gvsu.edu/speechlab/

▌**Research consultants:** The Center for Scholarly and Creative Excellence (CSCE) promotes a culture of active, engaged, ethical scholarship. It supports innovative faculty and student research and collaborative partnerships in the broader community. For more information, please visit their website: https://www.gvsu.edu/csce/

▌**Library:** GVSU's library offers a vast collection of online resources available for students. Visit their website for more details: https://www.gvsu.edu/library/

▌**Disability support resources:** If any student in this class has special needs because of a disability, please contact Disability Support Resources at http://www.gvsu.edu/dsr/ (DSR) at 616-331-2490.

Religious Observance: The university recognizes and respects religious traditions. If you require special accommodations for religious observances, inform the instructor in advance.

Emergency Procedures:

☞ **In Case of Emergency Fire**: Immediately proceed to the nearest exit during a fire alarm. Do not use elevators. More information is available on the University's Emergency website located at http://www.gvsu.edu/emergency.

Tentative Course Content (subject to change throughout the semester):

🖊 **January 20, Martin Luther King, Jr. Day Recess:** No classes!

🖊 **March 2-9 Spring Break:** No classes!

Week	Topics Covered
1	Introduction to data mining, tasks, and Python basics
2	Descriptive statistics, visualization, Numpy, and Pandas
3	Data cleaning, transformation, compression, and sampling
4	Similarity and distance measures
5	Feature relationships and dependencies
6	Midterm preparation and advanced transformation
7	Midterm exam
8	Feature extraction, selection, and Markov Blanket
9	Fall Break (No Class)
10	Decision trees
11	Model evaluation, selection, and Bayesian classification
12	Regression, perceptron, clustering, and lazy learning
13	Neural networks and CNNs
14	RNNs, attention, and transformers
15	Project presentation and final exam preparation
16	Final exam