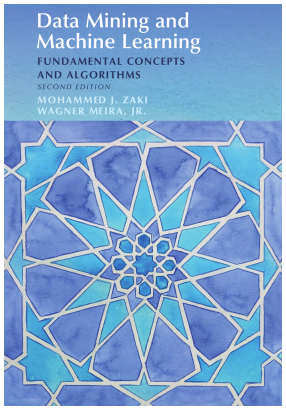
Course	**Knowledge Discovery from Data**
Lecture	MWF 11:10-12:00 PM
Lab	MWF 12:10-1:00 PM
Instructor	Professor Foaad Khosmood, Cal Poly CSSE
Office	14-204
Office Hours	MTWF 4:10-5:00
Email	foaad@calpoly.edu, (***must*** put "**[CSC 466]**" in subject line)
Home Page	https://canvas.calpoly.edu/courses/50163
Schedule	https://docs.google.com/spreadsheets/d/1pGOjPopxrAdyCPIJZs85sSuzJ9mepwcECu1Ev74PMfM/edit#gid=93074041
Slack	Everyone needs to join **#csc466-s22** on **foaadcp.slack.com** work space https://join.slack.com/t/foaadcp/shared_invite/zt-14jiqv0jm-xfP1ie3vo8hmlLa76lKbBw

Course Objectives

- Learn practical data processing and analysis skills
- Understand the theory and mathematical foundation of data mining processes
- Recognize different types of KDD procedures and identify their uses
- Interpret and analyze the results of KDD processes
- Recognize and evaluate societal impact of KDD technology, make informed choices about use of KDD technology



Course Texts

1. Zaki and Meira Jr., *Data Mining and Analysis, (2nd edition)* (required)
    a. You can read it online here, but I recommend buying it in a better form.
2. Various papers, online tutorials and manuals (provided by instructor)

Prerequisites

Prerequisite: CSC 349 and one of the following: STAT 302, STAT 312, STAT 321 or STAT 350.

Course Format

The course consists of lectures and labs. The lectures are where most material is presented. The labs are basically your own time to work on your lab/homework deliverable. Our quizzes will also be taken during lab times. Lab attendance is not mandatory. But do not schedule something else during lab times! I may ask you to come in for an important discussion, and you should be available. We make heavy use of Zoom, Canvas, Google Docs and Slack in this class. Please make sure you have access to all.

Labs

There will be a number of labs that are required for the course. The first few labs are designed to teach you actual skills, on various platforms, mostly about data processing and manipulation. The last few labs are about using those skills to solve larger problems. Unless otherwise stated, labs are individual works. No collaborations are allowed. You may know the material already, in which case it would be easy for you to finish the lab early. The labs all have the same number of points, but some of them will take you much longer than others. Do not make assumptions about the labs to come based on the content or duration of previous labs.

Quizzes

There will be at least four quizzes. These are designed to test:

1. Lecture material and readings that have been assigned for up until the day of the quiz
2. Skills and knowledge learned from previous labs and class discussions
3. Speed of execution for solving problems, and using the right tools

Quizzes for this class are open-book/notes/everything including the Internet, but not open classmates or any other person. You may not discuss, collaborate or seek the advice of anyone. They are strictly timed normally, however for the online delivery mode, I will allow up to about 12 hours to complete them. You can work on them anywhere and turn them in via Polylearn. Do not plan on any make up quizzes.

Projects

There will be 2-3 group projects and activities. The last one will be worth the most points and will have a presentation component. Projects will usually have multiple deliverable documents that may include data, code and writing.

Late Policy

Do not be late. It will not save you time, since the course is already very impacted. You will almost certainly miss an important part of the next assignment. But if you must do it, late assignments will receive a deduction of one full grade (10%) per 24 hour period, up to ***two*** such periods (this includes weekends and holidays). This policy may differ on some assignments but it is the official late policy unless otherwise stated.

Late days

From time to time, I may announce a relevant lecture or other extracurricular activities. These are not directly related to our class and will not be on quizzes or exams. But you can earn a free “late day” (i.e. no 10% deduction) if you participate. I will specifically announce these events in class and put them on Polylearn. Earning a late day means you do not get the penalty, but you do not get another day in return. No matter what, you can only be late 2 days maximum. Late days are strictly individual. Only project members who have the late day get the benefits.

Partnerships

On some of the work you may be working with partners. Collaboration tends to help with figuring out difficult concepts and generally makes the whole process more pleasant. A word of caution, though: While it is tempting to just divide up the work, be sure each partner understands the whole project. Concepts learned on the assignments could show up on quizzes which are worth far more than the individual assignments in the final analysis. Even if your partner bails you out of a tight spot, be sure you understand the work, or it will come back to haunt you. Only one copy of the assignment is necessary to turn in (unless otherwise stated) Be absolutely certain that all partners' names appear on all assignments. Credit will be given only to students whose names appear on the assignment.

Programming

This is not a programming class. The good news is I don’t care about your choice of editor, coding style or length of programs. Most labs specify the programming language to use. Even though we will be learning a number of new languages, you are free to use any language you like on your final project. The bad news is that I will not be able to debug any code for you. You will have to set up your own development environment, toolchain, debugging routine and process.

Slack

Slack is a great communication tool, and its use in this class is mandatory. Please make an account immediately if you haven’t done so already, and join the workspace and channel for this class. You must also check it daily. I highly, highly recommend the desktop client for Mac and Windows.

Online Code Repositories

This course involves group projects. The use of online code repositories like Github is very popular and efficient. You are free to use them subject to the following rules:

1. Under no circumstances can you use the work from student assignments of previous quarters, other courses, or other teams in the same course and term. You can only use a repository project set up this quarter with your own team members. If you are caught using code from another repository, it is considered academic dishonesty.
2. You must create private repositories only. After the current quarter ends, you must ensure your repository is deleted or remains private for the long term (5-7 years). There have been many cases of students using free private repositories only to have them turn public at some later point.

Submitting Written Work

Some of the homeworks will be problem sets from the text. These will be submitted using the Polylearn system. In general you will need to use **plain text,** paragraphs with a single line separating paragraphs from each other. Each response will begin with a problem number followed by “)” and a space. You do not need to repeat the problem statements.

Submitting Programs

Programming assignments will be submitted online via Polylearn. In general I will ask that all your code executes on the Computer Science Department servers Unix1-4 (also dev1), unless we specifically use another environment like Colab. When turning in programming assignments, be careful to submit your final version and to have tested it before submitting.

Grading

The grading breakdown is:

Graded item	Number	Value (total)
Labs	about 5-10	45%
Quizzes	4-5	20%
Activities/Projects	1	10%
Final Project	1	25%

Policy On Collaboration

Programming assignments in this class are intended to be demonstrations of individual or partnership abilities. To this end, programs are to be written only by the designated author(s). High-level discussion of problems and problem-solving techniques, however, is beneficial to all involved. You are encouraged to discuss approaches so long as those with whom you consult are given due credit in your program headers.

It is never acceptable to allow someone else to have your work for reference or to refer to someone else's work while writing your own.

In this case, "someone else's work" means not only other students' programs, but also materials from any other source, including, but not limited to, the world wide web, other reference books, or previous course materials. Collaboration that goes beyond general approaches or that is uncredited will be considered cheating. If you are unsure about what constitutes proper or improper collaboration, consult the instructor for guidance.

Policy On Cheating

Don't! I consider academic dishonesty a serious offense, especially on take-home quizzes and labs. Any instances of cheating or plagiarism will be referred to the Office of Student Rights and Responsibilities. The Cal Poly rules and policies are listed in the catalog, as well as at the OSRR web site, http://www.osrr.calpoly.edu. The general policy, however, is very simply stated in the Campus Administrative Manual (C.A.M. 684): ***Cheating requires an "F" course grade.***

Turning in work is presumed to be a claim of authorship unless explicitly stated otherwise. If the course rules are unclear or you are unsure of how they apply, ask the instructor beforehand.

Online Mode Policies and Tools

We **will not be using online mode** or zoom for this quarter until further notice. It's all in person. Lectures are not recorded. Please take good notes.

COVID-19

We will observe these COVID 19 protocols regardless of current campus regulations. Everyone needs to:

1. Wear your mask at all times indoors (Spring 2022 update: Yes, please continue to do this. I am old!)
2. Stay home if you have symptoms and get tested immediately
3. Notify me and the campus immediately if you have a positive test result
4. Do not eat or drink in class

Policy Changes

There might be changes to the stated policies in this syllabus, or other class rules during the quarter. There will very likely be changes to the class schedule and due dates. Most often these changes are to help students and loosen some restrictions. Changes will be announced in class (lecture) and on Polylearn.