

San José State University
Computer Engineering Department
CMPE255, Data Mining, Section 01, Spring, 2024

Course and Contact Information

Instructor(s): Kaikai Liu
Office Location: ENG257
Telephone: (408) (924-7847)
Email: kaikai.liu@sjsu.edu
Office Hours: See <http://www.sjsu.edu/cmpe/faculty/officehours.php>
Class Days/Time: Tuesday & Thursday 10:30AM-11:45AM
Classroom: ENG329
Prerequisites: CMPE 180A, classified standing, or instructor consent. Artificial Intelligence or Computer Engineering or Software Engineering majors only.

Course Description

Data representation and preprocessing, proximity, finding nearest neighbors, dimensionality reduction, exploratory analysis, association analysis and sequential patterns, supervised learning: model selection and evaluation, overfitting, clustering, advanced topics.

Course Format

Technology Intensive, Hybrid, and Online Courses

This course requires the student to have a personal computer that is installed with a modern operating system (Windows, MacOS, or Linux) and python development environment. Students will be asked to use their laptops or smart devices during the class, or offline, in order to participate in the class assignments, quizzes, and exams. Students must ensure that the laptop is in a good condition to be able to finish the assignments, take the quizzes and exams on Canvas, which may include Respondus monitoring using the laptop's camera. Please note that any missed or compromised assignments or exams resulting from technical laptop issues may not be rescheduled or retaken.

Faculty Web Page and MYSJSU Messaging

Course materials such as syllabus, handouts, notes, assignment instructions, etc. can be found on [Canvas Learning Management System course login website](#). You are responsible for regularly checking with the messaging system through [MySJSU](#) on [Spartan App Portal](#) (or other communication system as indicated by the instructor) to learn of any updates.

Course Goals

This course centers on data mining and its practical applications. We will comprehensively explore a variety of data mining algorithms and techniques, emphasizing their versatile utility in various domains. Topics span from predicting future figures to spam detection, text/image classification, and identifying objects in visual media. Our lectures will blend core concepts with case studies, facilitating a profound comprehension of the subject and its real-world applications.

The curriculum integrates both in-class and homework assignments to assess students' adeptness at deploying data mining effectively across diverse contexts. Through individual and group projects, students will gain hands-on experience, delving into intricate analyses of real-world data. These projects foster the development of essential engineering skills, encompassing teamwork, technical leadership, and articulate communication – both written and oral.

Course Learning Outcomes (CLO)

Upon successful completion of this course, students are expected to be able to:

- 1. CLO 1– Illustrate the fundamental concepts of the data mining process and its building blocks including algorithms, techniques, and applications.*
- 2. CLO 2 – Discuss and model the solution to a given problem using appropriate data mining algorithms, technologies, and tools.*
- 3. CLO 3 – Gain hands-on experience by performing an extensive analysis using state-of-the-art data mining tools, in individual and group projects.*
- 4. CLO 4 – Effectively present and communicate the knowledge they have acquired in the course through reports and oral presentations.*

Required Texts/Readings

Textbook

A designated textbook is not mandatory for this course. Lecture slides and curated readings from online resources comprehensively address the subject matter. Supplementary reference textbooks are also listed for students seeking additional context or deeper insights into the subjects covered.

Each student is accountable for referring to the updated syllabus on Canvas and the provided lecture slides. These resources will aid in identifying the readings aligned with the concepts taught each week.

Other Readings

Data Mining: Concepts and Techniques, by Jiawei Han, Micheline Kamber and Jian Pei, Morgan Kaufmann, Elsevier Inc. (available as eBook from the SJSU Library)

Introduction to Data Mining (Second Edition), by Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. Pearson Education India, 2016.

Data Mining: The Textbook, by Charu C. Aggarwal, Springer, May 2015 (Springer: <http://www.springer.com/us/book/9783319141411>)

Mining of Massive Datasets, by Jure Leskovec, Anand Rajaraman and Jeffrey Ullman, Cambridge University Press (download from <http://infolab.stanford.edu/~ullman/mmds/book.pdf>)

Other technology requirements / equipment / material

Students are required to possess a computer and necessary accessories to complete assignments and projects. Additionally, PDF software is essential for generating final reports. The class will utilize various programming languages, platforms, and software tools, all of which are available for free download.

Notably, Python will play a central role in enhancing class interactivity. Clear installation instructions will be accessible online. Python will be employed in numerous in-class activities, homework assignments, and projects.

Should students be more accustomed to using R or MATLAB over Python, they can seek special permission to utilize R or MATLAB for homework assignments.

Course Requirements and Assignments

Success in this course is based on the expectation that students will spend, for each unit of credit, a minimum of 45 hours over the length of the course (normally three hours per unit per week) for instruction, preparation/studying, or course related activities, including but not limited to internships, labs, and clinical practica. Other course structures will have equivalent workload expectations as described in the syllabus.”

Final Examination or Evaluation

The final exam will be closed book. Cheat sheet and electrical devices are not allowed. The culminating activity also include a final project report, software program, demo, and presentation.

Grading Information

- 5% Quiz
- 15% Midterm Exam
- 20% Homework and Assignment
- 30% Project
- 30% Final Exam

Each project necessitates the creation of a comprehensive design, project document, software programs, presentation file, video demo, and a final report. The project report should encompass detailed project information, encompassing the introduction, references, tutorials for key techniques, meticulous step-by-step setups, deliverables, and experimental outcomes.

Evaluation of the project will be based on the subsequent rubrics:

1. *Effective teamwork in achieving objectives (potential peer review may apply).*
2. *Relevance of the project topic to the course, with potential for serving as the culminating activity.*
3. *Demonstrated evidence of substantial engineering or research efforts within the deliverables.*
4. *The demonstration displaying strong potential for creating a minimum viable product (MVP) for the proposed application.*
5. *A well-organized and articulate report, largely devoid of grammatical errors and written with clarity.*

Determination of Grades

- *Final grades will be computed based on the cumulative weighted score of all assignments, including exams, homework assignments, and projects. The assignment of ABCDE grades will be determined by establishing thresholds that adhere to appropriate distribution and all requisite policies.*
 - *No opportunities for extra credit will be provided.*
 - *Peer ratings and comments may factor into project scores.*
 - *Reports must be submitted to Canvas by the specified deadline. Submissions will not be accepted once the Canvas assignment is closed.*
 - *For assignments enabled with Turnitin, submission formats must adhere to Turnitin's standards (such as WORD or PDF). Submissions failing to meet these criteria will be deemed late and subjected to corresponding penalties.*
 - *In select assignments, full marks (100%) are reserved for bonus points or "exceeding expectations". The points attributed to "meeting all requirements" will be outlined in each individual assignment.*
- Below are a few samples of grading scale. The point range for each letter grade is subject to change.*

<i>Grade</i>	<i>Points</i>	<i>Percentage</i>
<i>A plus</i>	<i>960 to 1000</i>	<i>96 to 100%</i>
<i>A</i>	<i>930 to 959</i>	<i>93 to 95%</i>
<i>A minus</i>	<i>900 to 929</i>	<i>90 to 92%</i>
<i>B plus</i>	<i>860 to 899</i>	<i>86 to 89 %</i>
<i>B</i>	<i>830 to 859</i>	<i>83 to 85%</i>
<i>B minus</i>	<i>800 to 829</i>	<i>80 to 82%</i>
<i>C plus</i>	<i>760 to 799</i>	<i>76 to 79%</i>
<i>C</i>	<i>730 to 759</i>	<i>73 to 75%</i>
<i>C minus</i>	<i>700 to 729</i>	<i>70 to 72%</i>
<i>D plus</i>	<i>660 to 699</i>	<i>66 to 69%</i>
<i>D</i>	<i>630 to 659</i>	<i>63 to 65%</i>
<i>D minus</i>	<i>600 to 629</i>	<i>60 to 62%</i>

Classroom Protocol

- 1. Students should attend all meetings of the class.*
- 2. Students are responsible for lecture, book sections, project presentations, and any instructions given in the class.*
- 3. Avoid disturbing the class: turn-off cell phones (or put them on vibrate mode), no messaging or chatting in the class or in the exams, no entering the class late or leaving early.*
- 4. Web browsing during the class is prohibited. Students are allowed to use computers for course related activities only. These activities include taking notes on the lecture underway, following the lecture on Web-based slides that the instructor has posted, and finding Web sites to which the instructor directs students at the time of the lecture.*
- 5. Students causing disruption in the class for other activities will be asked to leave the class and will be referred to the Judicial Affairs Officer of the University for disrupting the class after repeated offenses.*

University Policies

Per University Policy S16-9, relevant university policy concerning all courses, such as student responsibilities, academic integrity, accommodations, dropping and adding, consent for recording of class, etc. and available student services (e.g. learning assistance, counseling, and other resources) are listed on Syllabus Information web page (<https://www.sjsu.edu/curriculum/courses/syllabus-info.php>). Make sure to visit this page to review and be aware of these university policies and resources.

Department Policies

- Students who do not provide documentation of having satisfied the class prerequisite or corequisite requirements (if any) by the second class meeting will be dropped from the class.
- All non-proctored report (or similarly sized) assignments in courses where some of the final grade depends on prose writing will be submitted to Turnitin.com.
- Major exams in this class may be video recorded to ensure academic integrity. The recordings will only be viewed if there is an issue to be addressed. Under no circumstances will the recordings be publicly released.

CMPE255-01 / Data Mining, Spring, 2024 Course Schedule

The content outlined in this schedule is provisional and may be subject to modifications. Students are required to refer to the latest and comprehensive version of the schedule, which will be made available on Canvas.

Course Schedule

Week/Lesson /Module	Date	Topics, Readings, Assignments, Deadlines
1	1/25, 1/30	Introduction to Data mining, computing environment
2	2/1, 2/6	Linear Algebra review, Python data analytic tools, Exploratory Data Analysis
3	2/8, 2/13	Cloud Data Query and Analytics, Data visualization, dashboard, and data apps
4	2/15, 2/20	Probability review, data sampling and pre-processing, data modeling, supervised learning, regression
5	2/22, 2/27	Logistic regression, classification algorithms and model evaluation
6	2/29, 3/5	Data mining via Neural Network/Deep Learning
7	3/7, 3/12	Mining image data, feature extraction, CNNs
8	3/14, 3/19	Image mining and Deep Learning models
9	3/21, 3/26	Advanced image mining via Deep Learning models, Midterm Exam
10	3/28, 4/2 (Spring break)	Unsupervised learning, Dimensionality reduction
11	4/4 (Spring break), 4/9	Manifold learning, Clustering
12	4/11, 4/16	Time series and sequential data mining
13	4/18, 4/23	Text data preprocessing and mining, Embedding
14	4/25, 4/30	Text data mining via Transformer models
15	5/2, 5/7	Audio/speech data mining Project Demo and Presentations
16	5/9	Project Demo and Presentations
Final Exam		Thursday, May 16 9:45 AM-12:00 PM Check: https://www.sjsu.edu/classes/final-exam-schedule/spring-2024.php

9:45 AM-12:00 PM