CSC 177: Data Warehousing and Data Mining
(Online Synchronous. Tue/Thu self-directed team lab: on Tuesday)

California State University, Sacramento (CSUS), Fall 2023 Semester, 3 Credits

Class Times & Location:
Recorded videos are available for asynchronous learning.
Section **02: 5:30pm-6:45 pm, Online**

(Class on Tuesday: only self-directed team Lab).
ALL PRE-RECORDED VIDEO RECORDINGS WILL BE PROVIDED.
Asynchronous learning from video recordings is encouraged.
Tuesday: **Group/Team Labs.** Students meet in their teams to work together as a team on class projects. Instructor or TA will be present. Lab classes will be held in the classroom. Classroom attendance to work together as a team is highly encouraged, however attendance is optional, except on exam days. Exams will be administered on Canvas during scheduled exam day and time. For students who are working or interning remote, they may take special permission to write the exam online at the scheduled exam time.

Your instructor

Dr. Jagan Chidella

Office: **ONLINE ON ZOOM**

Office Hours:
Fri: 4-5pm (by appointment in advance recommended)

Email: jagannadha.chidella@csus.edu

Phone: (916)316-8506

Dr Chidella teaches undergrad level and grad level courses in Computer Science at CSUS. In the last 30 years he has designed and implemented systems for US Army, Oracle/Sun Microsystems, Hewlett Packard, and three State of California Agencies. He has also done AI and Data Mining research and developed frameworks for Center for AI and Robotics (CAIR, Bangalore), Carnegie-Group Inc/CenturyLink Telecom, US Army, Xerox Corporation, Spencer-Trask and was a founding member at three start-up AI companies. His doctoral dissertation is on a framework with constraint solvers to transform and interpret data and extract knowledge.

Email Policy

Please email me directly or use the Canvas messaging system. Please check your Sac State email at least once a day, in case I reach out to you through email, via Canvas announcements.

Course Content

This course is on data warehousing and data mining. Data warehousing involves data preprocessing, data integration, and providing on-line analytical processing (OLAP) tools for the interactive analysis of multidimensional data, which facilitates effective data mining. Data mining is the automated extraction of hidden predictive information from databases. Data mining applies concepts and techniques from the fields of databases, machine learning, algorithms, information retrieval, and statistics. Topics include data warehousing, association analysis, classification, clustering, numeric prediction, and selected advanced data mining topics. There are two aspects to this task: concept learning and practical projects. We will survey both aspects and apply them by building systems in teams of up to 4 members during the semester.

Prerequisites

CSC 134 and STAT 50 or ENGR 115.

Goals of the Course

The overall objective of this course is surveying the field of data warehousing and data mining and applying the knowledge in team projects using the Python programming language. More specifically, by the end of this course you will be able to:

1. Understand data concepts, data preprocessing and simple linear regression and implement a project.
2. Understand classification decision trees and implement a project.
3. Understand different classification methods and algorithms and implement a project.
4. Understand clustering techniques and implement a project.
5. Understand Association Analysis and implement a project
5. Understand special topics such as data integration and outlier analysis.

Required Texts

The primary textbook for this course is Introduction to Data Mining (second edition, Pearson Publishing) by Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar.

The textbook is not mandatory. The PowerPoint slides material will be available.

Attendance and Participation

Attendance is not required except we will have labs and demos in class. However, attendance is expected in the sense that material missed because of unexcused absences will not be provided by the instructor on other occasions. In other words, there will be no private lectures during office hours for student who do not make it to class, unless there is a critical reason. All students are expected to participate in their groups during in-class activities and during class discussions. There are no participation grades, however.

Methods of Evaluation

Assignments	**Weight**
Team Projects: Code Implementation	40%
Team Projects: Documentation and Presentations	10%
Total Team Projects	50%
Final Exam	25%
Mid Term	25%
Total Exams	50%
Quizzes (four)	3% (used for bonus)
Bonus: Quizzes/Special Presentations/Papers	3% bonus
Total	100% + 3% bonus

At the end of the semester, a final percentage will be calculated according to the above criteria. It will then be rounded to the nearest integer value. Then, a letter grade according to the following scale will

be assigned. (**No curving will be further performed**.). However, extraordinary performance in any work assigned will be used to reward a student in border cases.

Range	Letter Grade
93-100	A
90-92	A-
87-89	B+
83-86	B
80-82	B-
77-79	C+

Range	Letter Grade
73-76	C
70-72	C-
67-69	D+
63-66	D
60-62	D-
59 or Less	F

Labs

Attendance is required. Please check Canvas "Modules" for lab dates. Labs will be presented by Instructor or TA.

Quizzes

There will be frequent preparatory Quizzes every two weeks which will be evaluated. The percentage of quizzes is minimal as the goal is to help students prepare for the Final. It is the student's responsibility to evaluate their performance in the quizzes and adjust their learning in preparation for the Final.

Team Project

There will be three or more Team Projects (on about six topics). Late submissions are usually accepted within TWO days after deadline, with 20% off penalty per day, unless otherwise noted.

Important Note: **Quality documentation is critical for the success of the project and counts toward your project grade. All work submitted must include documentation.**

Exams

There will be a final exam during the finals week.

Missed and Late Assignment Policy

If you are unable to take an exam at the scheduled time because of illness or other problems, you must contact me **beforehand** to arrange to take the exam at a different time. Failure to make prior arrangements for a missed exam will result in a grade of 0 for the exam.
In-class work missed because of absence will only be accepted if arrangements are made **beforehand**.
Late projects will be accepted within 2 days after due dates, with 20% penalty for each day. Alternate due dates can be arranged in special circumstances provided these arrangements are made **before** the due date.

Tentative Schedule

The following schedule is a tentative plan. Lectures may not be in the same order. Note that November 10th (Veterans Day) and November 23rd, and November 24th (Thanksgiving) are holidays and campus is closed. The Fall 2023 Class sessions will end December 8th and the final exam will be held in the Finals Week of December 11-15, 2023.

WEEK	Tentative Lecture Topic	Quiz	Projects due
1. 08/29 Tuesday (lab) 08/31 Thursday	Introduction. Python Coding		
2. 09/05 Tuesday (lab) 09/07 Thursday 09/04 MONDAY LABOR DAY HOLIDAY	Understanding Data Python Coding		
3. 09/12 Tuesday (lab) 09/14 Thursday	Association Rule Analysis Python Coding		
4. 09/19 Tuesday (lab) 09/21 Thursday	Association Rule Analysis Python Coding		Association Analysis Project
5. 09/26 Tuesday (lab) 09/28 Thursday	Dimensionality Reduction, Backward Elimination & other Data Preprocessing		Association Analysis presentations
6. 10/03 Tuesday (lab) 10/05 Thursday	Simple Linear Regression + Lasso & Ridge Regression Python Coding		Data Preprocessing & Linear Regression Project
7. 10/10 Tuesday (lab) 10/13 Thursday	Decision Trees Python Coding		DP & LP Presentations
8. 10/17 Tuesday (lab) 10/19 Thursday	Model Overfitting & Underfitting decision trees Python Coding		Decision Tree Project due
9. 10/24 Tuesday (lab) 10/26 Thursday	Other classification techniques continued (SVM, Logistic, Bayesian, ANN) Python Coding	Mid Term	Decision Trees Presentations
10. 10/31 Tuesday (lab) 11/02 Thursday	Other classification techniques continued		Classification Project

	(SVM, Logistic, Bayesian, ANN) Python Coding		
11. 11/07 Tuesday (lab) 11/09 Thursday NOVEMBER 10 VETERAN's HOLIDAY	Clustering Techniques Python Coding	15-minute Quiz-3 (Take Home)	Classification Presentations
12. 11/14 Tuesday (lab) 11/16 Thursday	Clustering Techniques Python Coding		Clustering Project
13. 11/21 Tuesday (lab) 11/23 Thursday 11/23-11/24 are THANKSGIVING HOLIDAYS	OPTIONAL: Anomaly Detection/Outlier Analysis Python Coding\	NO recordings available on Anomaly Detection. I may create new recordings.	Clustering Project presentations
14. 11/28 Tuesday (lab) 11/30 Thursday	OPTIONAL: Anomaly Detect/Outlier Analysis Python Coding	15-minute Quiz-4 (Take Home)	Anomaly Detection Project
15. 12/5 Tuesday (lab) 12/7 Thursday	Special Topics, Presentations		Project presentations
16. FINAL EXAMS WEEK	December 11-15		

University Policies

Academic Honesty

If you violate the University's Honor Code (https://www.csus.edu/umanual/student/stu-0100.htm), you will receive a reduced or failing grade in the course, other penalties may be imposed, and the violation will be reported to the Student Conduct Officer. Automated tools may be used on any assignment, at any time, to detect inappropriate collaboration and to determine the originality of submissions.

Adding/Dropping

You are responsible for enrolling in courses and verifying your schedule on MySacState. Please refer to the Fall 2022 Calendar in https://catalog.csus.edu/academic-calendar/#fall2022text

I do not give "Incomplete" grades to students requesting a drop after the deadline except in extraordinary circumstances.

Disability Services

If you have a documented disability and need accommodations in this course, please register with the Office of Services to Students with Disabilities (https://www.csus.edu/sswd/). They will verify your need for services and make recommendations for the course. I will be happy to discuss any accommodations I can provide to assist your learning with you.

Religious Observation Accommodations

If you cannot satisfy a requirement of the course for religious reasons you must let me know at least two weeks in advance. In some cases, you will be required to make up the requirement; in other cases the requirement may be waived with suitable adjustment in grading criteria.

Excused Absences

Students who are unable to attend class due to Sac State sponsored activities (such as sports, band, academic competition, field trips, etc.) or personal religious observances may request reasonable accommodations. Please notify me during the first week of class regarding potential absences so that we can determine alternative methods for you to complete the required work.

Housing & Food Security

If you experience difficulties with financial, housing or food security, please contact Basic Needs Division of Student Affairs (https://www.csus.edu/basicneeds/) for assistance.

Parents & Families

If you are students with children, please feel free to let me know your needs. Also, please reach out to Parents & Families Division of Student Affairs (https://www.csus.edu/student/parents/student-parents/) for all resources available on campus.

Changes to this Document

I reserve the right to change any information on this document or course materials at any time.