

## **CSC 177: Data Warehousing and Data Mining**

California State University, Sacramento (CSUS), Summer 2019 Semester, 3 Credits

### **Class Times & Locations:**

Section 01: Tu/Th 5:30-9 pm, Room RVR 1010

### **Your Instructor**

Dr. Jagan Chidella



**Office:** Riverside Hall 5004

#### **Office Hours:**

Tu/Th : 4:15-5:15pm

**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 5-6 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 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	50%
Team Project Documentation and Presentations	20%
Final Exam	30%
Original Paper (Bonus)	10% * (bonus)
Preparatory 15-minute Quizzes	0%
Total	100% + 10% 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 four preparatory Quizzes which will be evaluated. However, the quizzes will not be included in the grade. 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. 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 on the last day of regular classes. There is no final exam week in a summer session.

## 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. Note that Thursday July 4<sup>th</sup> is Independence Day holiday, campus is closed, and no classes held. Hence the Summer Session will end on Tuesday July 2<sup>nd</sup>. There is no separate final exam week in a summer session. The final exam will be held on July 2<sup>nd</sup>.

WEEK	Tentative Lecture Topic or Lab	Quiz	Projects
1. 05/28 Tuesday	-Introduction. -Data Preprocessing		
2. 05/30 Thursday	-Data Preprocessing. -Linear Regression		
3. 06/04 Tuesday	Decision Tree classification	15-minute Quiz-1	
4. 06/06 Thursday	Decision Tree classification		Data Preprocessing Project due
5. 06/11 Tuesday	Other classification techniques (SVM, Logistic, Bayesian, ANN)	15-minute Quiz-2	
6. 06/13 Thursday	Other classification techniques continued (SVM, Logistic, Bayesian, ANN)		Decision Tree Project due
7. 06/18 Tuesday	Clustering Techniques	15-minute Quiz-3	
8. 06/20 Thursday	Extra Topics		Classification Techniques Project due
9. 06/25 Tuesday	Extra Topics	15-minute Quiz-4	
10. 06/27 Thursday	Project Team Presentations		Clustering Techniques Project due
11. 07/02 Tuesday	FINAL EXAM		
12. 07/04 Thursday	Independence Day holiday CAMPUS CLOSED -NO CLASSES		

# **University Policies**

## **Academic Honesty**

If you violate the University's Honor Code (<https://www.csus.edu/umannual/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 Spring 2019 Calendar in <http://catalog.csus.edu/academic-calendar/#spring2019text> . 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.