



CALIFORNIA STATE UNIVERSITY, CHICO

CSCI 598: DATA MINING

Abbreviated Syllabus for Fall Semester 2006

Visit <http://www.ecst.csuchico.edu/~juliano/csci598> for additional detail.

Prerequisites

- CSCI 311, *Algorithms and Data Structures*; or
- Graduate standing

Description

3 units. This course introduces the student to basic concepts, tasks, methods, and techniques in data mining; in particular, the course focuses on practical machine learning tools and techniques used in data mining. Students will develop an understanding of the data mining process and issues, learn various techniques for data mining, and apply the techniques in solving data mining problems using data mining tools and systems.

“Data mining, also known as knowledge-discovery in databases (KDD), is the practice of automatically searching large stores of data for patterns. To do this, data mining uses computational techniques from statistics and pattern recognition.”

- From http://en.wikipedia.org/wiki/Data_mining

Note: Students from departments such as Statistics, Biology, Mathematics, and Electrical & Computer Engineering who are working in interdisciplinary research (e.g., bioinformatics, modeling, data analysis) are especially encouraged to take this course.

Class #	Section	Act	Days	Time	Room	Instructors
5900	CSCI 598-02	DIS	TR	1100-1215pm	MLIB 031	Dr. J Juliano@csuchico.edu

Instructor Information

Dr. Juliano (a.k.a. Dr. J)
<http://www.ecst.csuchico.edu/~juliano>

Office Hours: M 2-5pm and W 3-5pm
OCNL 222
Tel 530 898-4619 / 6442 (dept)
Fax 530 898-5995
Appointments / walk-ins welcome.

Required Textbook

Data Mining: Practical Machine Learning Tools and Techniques, 2/e
Ian Witten and Eibe Frank, 2005.
Elsevier Inc. Burlington, Massachusetts.
ISBN 0-12-088407-0

Additional Requirements

1. Students are expected to open and maintain a *Chico State Connection* (CSC) Portal (see <http://portal.csuchico.edu>) account in order to access up-to-date *WebCT* tools that include an on-line calendar of events, current scores, discussion board, etc.
2. Students will also be using **WEKA**, the **Waikato Environment for Knowledge Analysis**, an open

source data mining software application written in Java and developed at the University of Waikato in New Zealand.

3. Students are expected to familiarize themselves with Dr. J's general policies and expectations detailed at [/~juliano/Teaching/Policies.html](http://www.ecst.csuchico.edu/~juliano/Teaching/Policies.html)

Grade Evaluation

30%	Written homework or assignments
30%	Laboratory (<i>WEKA</i>) projects
35%	Research paper presentation (oral & written)
5%	Class participation (local students)

Also see the on-line syllabus for details of final grade calculation.

Selected Topics for Research Papers:

Web search and Web mining, data mining for fraud detection, mining text and sequential data, data mining in bioinformatics (genomics, proteomics, etc).

Additional Information

<http://www.ecst.csuchico.edu/~juliano/csci598/>
<http://portal.csuchico.edu/>