

NEW COURSE PROPOSAL

Courses must be submitted by October 15, 2013, and finalized by the end of that fall semester for the next catalog production. (Late submitssion FOR 2015-2016 Catalog)
Use YELLOWED areas to enter data.

DATE (*Change if modified and redate file with current date*)

4/21/14, 9/16/14

PROGRAM AREA(S)

COMPUTER SCIENCE

1. Course Information. *[Follow accepted catalog format.]***Prefix(es)** (Add additional prefixes if cross-listed) and **Course No.** COMP 478**Title:** INTRODUCTION TO DATA MINING **Units:** 3

x Prerequisites MATH151, MATH240, MATH352, COMP350, COMP362

Corequisites

x Consent of Instructor Required for Enrollment

Catalog Description (Do not use any symbols): **Practical programming introduction to tools, methods, techniques, and workflows for understanding large quantities of data. Familiarizes students with instruments for data representation, preprocessing including data cleaning and reduction, input feature selection, data analysis including classification, clustering, and prediction, and data visualization.**

Grading Scheme:

x A-F Grades

Credit/No Credit

Optional (Student Choice)

Repeatability:

Repeatable for a maximum of units

Total Completions Allowed

Multiple Enrollment in Same Semester

Course Level Information:

x Undergraduate

Post-Baccalaureate/Credential

Graduate

Mode of Instruction/Components *(Hours per Unit are defaulted).*

	Units	Hours per Unit	Default Section Size	Graded Component	CS & HEGIS # (Filled in by the Provost's Office)
Lecture	2	1	24	x	
Seminar		1			
Laboratory	1	3	24	x	
Activity		2			
Field Studies					
Indep Study					
Other Blank					

Leave the following hours per week areas blank. The hours per week will be filled out for you.

hours lecture per week

hours blank per week

2. Course Attributes:

General Education Categories: All courses with GE category notations (including deletions) must be submitted to the GE website: <http://summit.csuci.edu/geapproval>. Upon completion, the GE Committee will forward your documents to the Curriculum Committee for further processing.

A (English Language, Communication, Critical Thinking)

A-1 Oral Communication

A-2 English Writing

A-3 Critical Thinking

B (Mathematics, Sciences & Technology)

B-1 Physical Sciences

B-2 Life Sciences – Biology

B-3 Mathematics – Mathematics and Applications

B-4 Computers and Information Technology

C (Fine Arts, Literature, Languages & Cultures)

C-1 Art

C-2 Literature Courses

C-3a Language

C-3b Multicultural

D (Social Perspectives)

E (Human Psychological and Physiological Perspectives)

UDIGE/INTD Interdisciplinary

Meets University Writing Requirement (Graduation Writing Assessment Requirement)

Meets University Language Requirement

American Institutions, Title V Section 40404: ☐ Government ☐ US Constitution ☐ US History

Regarding Exec Order 405, for more information: <http://senate.csuci.edu/comm/curriculum/resources.htm>

Service Learning Course (Approval from the Center for Community Engagement must be received before you can request this course attribute).

Online Course (Answer YES if the course is ALWAYS delivered online).

Lab Fee Request – Lab fee requests should be directed to the Student Fee Committee.

3. Justification and Requirements for the Course. (Make a brief statement to justify the need for the course)

A. Justification: Data and algorithms have always been in the center of the core computer science education. However, the recent dramatic increase in the volume of data in our daily lives requires designing special algorithms, tools, and programming techniques that will make it possible to analyze that flood of information to the benefits of the society. Every aspect of our lives is affected by the growth, and virtually every institution needs to develop expertise in managing its data. Properly prepared computer scientists provide that kind of capability since they possess the necessary rigorous mathematical knowledge augmented by programming skills that turn that knowledge into practical use.

B. Degree Requirement:

☐ Requirement for the Major/Minor

☒ Elective for the Major/Minor

☐ Free Elective

Note: Submit Program Modification if this course changes your program.

4. Student Learning Outcomes. List in numerical order. Please refer to the Curriculum Committee's "Learning Outcomes" guideline for measurable outcomes that reflect elements of Bloom's Taxonomy:

<http://senate.csuci.edu/comm/curriculum/resources.htm>. The committee recommends 4 to 8 student learning outcomes, unless governed by an external agency (e.g., Nursing).

Upon completion of the course, the student will be able to:

- Select data representation suitable for analysis
- Preprocess the data for analysis by, for example, removing noise, normalization
- Select input representation in a form of a feature vector
- Design data processing workflows
- Implement techniques for multilevel and multiview analysis
- Evaluate and compare analytical models using cross-validation techniques
- Refine the data analysis by addressing bias and variance in the data
- **Discuss** ~~Discourse~~ problems in data mining at multiple levels

5. Course Content in Outline Form. [Be as brief as possible, but use as much space as necessary]

- Overview of Data Mining
- Obtaining, cleaning, and preprocessing data
- Feature selection and engineering
- Dimensionality reduction
- Programming linear regression models
- Programming logistic regressions models
- Programming classification models
- Programming clustering models
- Bias-variance tradeoffs
- Cross-validation techniques

Does this course content overlap with a course offered in your academic program? **Yes** ☐ **No** ☒

If YES, what course(s) and provide a justification of the overlap.

Does this course content overlap a course offered in another academic area? **Yes** ☐ **No** ☒

If YES, what course(s) and provide a justification of the overlap.

Overlapping courses require Chairs' signatures.

6. Cross-listed Courses (Please note each prefix in item No. 1) Beyond three disciplines consult with the Curriculum Committee.

A. List Cross-listed Courses (Signature of Academic Chair(s) of the other academic area(s) is required).

List each cross-listed prefix for the course: N/A

B. Program responsible for staffing: Computer Science

7. References. [Provide 3 - 5 references]

1. *Practical Machine Learning Tools and Techniques 3rd Edition*, Ian H. Witten Eibe Frank Mark A. Hall, Morgan Kaufmann Publishers, Elsevier 2011
2. *Data Mining: Concepts and Techniques 2nd Edition*, Jiawei Han, Micheline Kamber, Morgan Kaufmann Publishers, Elsevier 2006
3. *Introduction to Machine Learning 2nd Edition*, Ethem Alpaydın, The MIT Press, 2010
4. *Python for Data Analysis*, Wes McKinney, O'Reilly Media, 2013
5. *Building Machine Learning Systems with Python*, Willi Richer, Luis Pedro Coelho, Packt Publishing, 2013

8. Tenure Track Faculty Qualified to Teach This Course.

Any CS Faculty

9. Requested Effective Date:

First semester offered: Fall 2015

10. New Resources Requested. **Yes** ☐ **No** ☒

If YES, list the resources needed.

A. Computer Needs (data processing, audio visual, broadcasting, other equipment, etc.)

Regular CS computer lab (upper-level computer power for processing and visualization of large amount of data)

B. Library Needs (streaming media, video hosting, databases, exhibit space, etc.)

N/A

C. Facility/Space/Transportation Needs

N/A

D. Lab Fee Requested **Yes** ☐ **No** ☒ (**Lab fee requests should be directed to the Student Fee Committee**)

E. Other

N/A

11. Will this new course alter any degree, credential, certificate, or minor in your program? **Yes** ☐ **No** ☒

If, YES attach a program update or program modification form for all programs affected.

Priority deadline for New Minors and Programs: **October 1, 2013** of preceding year.

Priority deadline for Course Proposals and Modifications: **October 15, 2013**, of preceding year.

Last day to submit forms to be considered during the current academic year: **April 15th**.

AJ Bieszczad

2/2414

Proposer of Course (Type in name. Signatures will be collected after Curriculum approval)

Date

Approval Sheet

Program/Course: Computer Science / COMP478 Introduction to **Data Mining** Machine Learning

If your course has a General Education Component or involves Center affiliation, the Center will also sign off during the approval process.

Multiple Chair fields are available for cross-listed courses.

The CI program review process includes a report from the respective department/program on its progress toward accessibility requirement compliance. By signing below, I acknowledge the importance of incorporating accessibility in course design.

Program Chair		
Signature		Date
Program Chair		
Signature		Date
Program Chair		
Signature		Date
General Education Chair		
Signature		Date
Center for International Affairs Director		
Signature		Date
Center for Integrative Studies Director		
Signature		Date
Center for Multicultural Engagement Director		
Signature		Date
Center for Civic Engagement Director		
Signature		Date
Curriculum Chair		
Signature		Date
AVP		
Signature		Date