



# Course Approval Form

For approval of new courses and deletions or modifications to an existing course.

registrar.gmu.edu/facultystaff/curriculum

### Action Requested:

Create new course       Delete existing course

Modify existing course (check all that apply)

Title       Credits       Repeat Status       Grade Type

Prereq/coreq       Schedule Type       Restrictions

Other: Number \_\_\_\_\_

### Course Level:

Undergraduate \_\_\_\_\_

Graduate \_\_\_\_\_

College/School:  Department:

Submitted by:  Ext:  Email:

Subject Code:  Number:  Effective Term:  Fall  Spring  Summer

(Do not list multiple codes or numbers. Each course proposal must have a separate form.)

Year

Title: Current

Banner (30 characters max including spaces)

New \_\_\_\_\_

Credits: (check one)  Fixed  Variable

or

Repeat Status: (check one)  Not Repeatable (NR)  Repeatable within degree (RD)  Repeatable within term (RT)

Maximum credits allowed:

Grade Mode: (check one)  Regular (A, B, C, etc.)  Satisfactory/No Credit  Special (A, B C, etc. +IP)

Schedule Type Code(s): (check all that apply)  Lecture (LEC)  Lab (LAB)  Recitation (RCT)  Internship (INT)

Independent Study (IND)  Seminar (SEM)  Studio (STU)

Prerequisite(s):

Corequisite(s): \_\_\_\_\_

Instructional Mode:

100% face-to-face

Hybrid: ≤ 50% electronically delivered

100% electronically delivered

Special Instructions: (list restrictions for major, college, or degree; hard-coding; etc.) \_\_\_\_\_

Are there equivalent course(s)?  Yes  No

If yes, please list \_\_\_\_\_

### Catalog Copy for NEW Courses Only (Consult University Catalog for models)

<b>Description</b> (No more than 60 words, use verb phrases and present tense)	<b>Notes</b> (List additional information for the course)
Data mining techniques from statistics, machine learning, and visualization to scientific knowledge discovery. Students will be given a set of case studies and projects to test their understanding of this field and provide a foundation for future applications in their careers.	
Indicate number of contact hours: Hours of Lecture or Seminar per week: <input type="text" value="3"/> Hours of Lab or Studio: <input type="text" value="0"/>	
When Offered: (check all that apply) <input checked="" type="checkbox"/> Fall <input type="checkbox"/> Summer <input checked="" type="checkbox"/> Spring	

Approval Signatures

  Date

Department Approval \_\_\_\_\_ College/School Approval \_\_\_\_\_ Date \_\_\_\_\_

If this course includes subject matter currently dealt with by any other units, the originating department must circulate this proposal for review by those units and obtain the necessary signatures prior to submission. Failure to do so will delay action on this proposal.

Unit Name	Unit Approval Name	Unit Approver's Signature	Date

**For Graduate Courses Only**

\_\_\_\_\_  
Graduate Council Member

\_\_\_\_\_  
Provost Office

\_\_\_\_\_  
Graduate Council Approval Date

**For Registrar Office's Use Only:** Banner \_\_\_\_\_ Catalog \_\_\_\_\_ *revised 2/2/10*

**Course Proposal Submitted to the Curriculum Committee of the College of Science**

**1. COURSE NUMBER AND TITLE:**

CDS 303: Scientific Data Mining

**Course Prerequisites:** CDS 101 or CDS 130 or equivalent, or permission of instructor

**Catalog Description:** Data mining techniques from statistics, machine learning, and visualization to scientific knowledge discovery. Students will be given a set of case studies and projects to test their understanding of this field and provide a foundation for future applications in their careers.

**2. COURSE JUSTIFICATION:** This is the same course as the previously approved CDS 401, but with the easing of prerequisites, therefore making the lower course number appropriate.

**Course Objectives:**

1. to develop an understanding of data mining and its scientific applications;
2. to become familiar with a variety of data mining concepts, techniques, and algorithms;
3. to become capable in applying these techniques and algorithms to solve scientific problems;
4. to provide a foundation and develop the skills for future data-intensive applications.

**Course Necessity:** Already approved course.

**Course Relationship to Existing Programs:** Already approved course.

**Course Relationship to Existing Courses:** Already approved course.

**3. APPROVAL HISTORY:** CDS 401 was approved in December 2006.

**4. SCHEDULING AND PROPOSED INSTRUCTORS:**

**Semester of Initial Offering:** Spring 2014

**Proposed Instructors:** Borne, Wegman

**5. TENTATIVE SYLLABUS:** See attached.

## CDS 401 Scientific Data Mining Fall Semester 2012



- **\*\*\*\*IMPORTANT NOTE:** Last Day to Add Classes = September 4, 2012.
  - **Course Syllabus Website:** <http://classweb.gmu.edu/kborne/cds401/>
  - **Supplemental Syllabus Information:**
    - Disability Services, Counseling Services, Other Campus Resources, University Policies, and more: <http://classweb.gmu.edu/kborne/supplemental-syllabus-information.htm>
    - Academic Skills Workshops - Academic Year 2012-2013: <http://caps.gmu.edu/learningservices/workshops.php>
    - Hours of Operation for Campus Offices and Services - Fall 2012: <http://www.gmu.edu/resources/welcome/fallhours.html>
    - Academic Calendars (Semester Calendar, Final Exam Schedule): <http://registrar.gmu.edu/calendars/>
  - **Honor Code:**
    - Instructors may submit Exam Papers, Homework solutions, or any other student assignment to either the [TurnItIn.com](http://turnitin.com) or the [SafeAssign](http://safeassign.com) plagiarism-detection services, in compliance with all of the following: GMU policy, Provost approval, and the [GMU Honor Code](#).
    - [Plagiarism](#) will not be tolerated.
- 
- **TO BE UPDATED: Reading Assignments:** <http://classweb.gmu.edu/kborne/cds401/cds401-reading.htm>
  - **Online Course Material:** Please log into <http://mymason.gmu.edu/> each week to get announcements, lecture slides, assignments, and grades.
  - **Supplemental Files and Related Documents:** <http://classweb.gmu.edu/kborne/cds401/files/>
  - **Lecture Day/Time:** Tuesdays and Thursdays 1:30-2:45 PM (August 28 – December 6) (see <https://patriotweb.gmu.edu/>)
    - **No Class on Tuesday October 9 or on Thursday November 22.**
    - **OTHER DATES WHEN CLASS DOES NOT MEET: TO BE ANNOUNCED.**
  - **Lecture Place:** Innovation Hall, Room 139 ([Fairfax campus](#))
  - **Midterm Exam:** Tuesday October 16, 2012 at 1:30pm (in our classroom)
  - **Final Exam:** [Tuesday December 18, 1:30pm-4:15pm](#) (in our classroom)
  - **Grading:**
    - 30% = Homework and Lab Exercises (the lowest 2 scores will be dropped)
    - 10% = Class Participation
    - 20% = Midterm Exam
    - 40% = Final Exam
  - **Course Instructor:** [Dr. Kirk Borne](#), Professor of Astrophysics and Computational Science
    - Office: [Research Hall](#), Room 357, phone 703-993-8402 (with voicemail)
    - Office Hours: **by appointment**
    - Instructor's Travel Schedule: <http://classweb.gmu.edu/kborne/travel-schedule.htm>
    - E-Mail: [kborne\(at\)gmu\(dot\)edu](mailto:kborne(at)gmu(dot)edu) (Students: <http://masonlive.gmu.edu/>) (Faculty: <https://mail.gmu.edu/>)
    - Mailbox: Mailstop 6A2, Research Hall.

- **Required Textbook:**
  - P.-N. Tan, M. Steinbach, & V. Kumar, [\*Introduction to Data Mining\*](#). Addison-Wesley, 2005. ISBN: 9780321321367.
- **Optional Supplemental Reading:**
  - M. Dunham, [\*Data Mining: Introductory and Advanced Topics\*](#). Prentice-Hall, 2002. ISBN: 9780130888921.
  - R. J. Roiger & M. W. Geatz, [\*Data Mining: A Tutorial-Based Primer\*](#). Addison-Wesley, 2002. ISBN: 9780201741285.
- **Technology Requirements:**
  - Access to Internet. Active user accounts on [myMason.gmu.edu](http://myMason.gmu.edu) and on <http://masonlive.gmu.edu/>.
  - Laptop (or equivalent portable device), to use during class exercises.
- **Course Description (from [GMU course catalog](#)):**
  - **(3 credits) Data mining techniques from statistics, machine learning, and visualization as applied to scientific knowledge discovery. Students will be given a set of case studies and projects to test their understanding of this field and to provide a foundation for future applications in their careers.**
- **Detailed Course Overview:**
  - This course provides a broad overview of the data mining component of the knowledge discovery process, as applied to scientific research. Scientific databases are growing at near-exponential rates. As the amount of data has grown, so has the difficulty in analyzing these large databases. Data mining is the search for hidden, meaningful patterns in such databases. Identifying these patterns and rules can provide significant competitive advantage to scientific research projects and in other career settings. Data mining is motivated and analyzed as the killer app for large scientific databases. Data mining techniques, algorithms, and applications are covered, as well as the key concepts of machine learning, data types, data preparation, previewing, noise handling, feature selection, normalization, data transformation, similarity measures, and distance metrics. Algorithms and techniques will be analyzed specifically in terms of their application to solving particular problems. Several scientific case studies will be presented from the science research literature. The techniques that are presented will be drawn from well known statistical, machine learning, visualization, and database algorithms, including clustering, decision trees, regression, Bayes theorem, nearest neighbor, neural networks, and genetic algorithms. Topics will include informatics, semantic knowledge mining, and the integration of data mining with large (and often distributed) scientific databases.
- **Prerequisites:**
  - [CDS 302](#) (Scientific Data and Databases).
  - Or else ... Permission of Instructor, depending on whether you are a Junior or Senior majoring in science
- **Course Objectives:**
  - to develop an understanding of data mining and its scientific applications;
  - to become familiar with a variety of data mining concepts, techniques, and algorithms;
  - to become capable in applying these techniques and algorithms to solve scientific problems; and
  - to provide a foundation and develop the skills for future data-intensive applications in the student's career.

---

Author: [Kirk D. Borne](#)

Last Update: 28 August 2012