

ERS 602: DATA MINING IN EARTH SCIENCES

Sudarshan S. Chawathe

Andrei V. Kurbatov

University of Maine

Fall 2016 *preliminary version*

This course studies the application of data mining and related computational tools to problems in the Earth Sciences. The approach is data-centric, with attention devoted to all aspects of data management, ranging from raw data acquisition to advanced analysis and visualization. However, the use of data mining techniques is of particular interest. The course is most suitable to students in the Earth Sciences who wish to apply modern tools and techniques from computational fields such as big data, data mining, parallel processing to datasets from their research areas. Further details will be determined early in the semester based on student backgrounds and interests.

Prerequisite: Permission of the instructors.

News and Reminders:

- Please read the class newsgroup for timely announcements: [umaine.ers602](mailto:umaine.ers602@creak.um.maine.edu) on NNTP server creak.um.maine.edu. Web interface to get started: <http://cs.umaine.edu/~chaw/news/>.
- The most recent version of this document may be found at <http://cs.umaine.edu/~chaw/ers602/>.
- Please use the PDF version of this document for printing and reference: [ers602.pdf](#)

Contact Information

Class meetings:

Time: Tuesdays 9:00–11:00 a.m.

Location: Sawyer Environmental Research Center, Room 108.

Instructors:

- Sudarshan S. Chawathe

Office: Neville Hall, Room 224.

Office hours: (Please check for changes.)

Tuesdays and Thursdays: 3:15–4:30 p.m.

Phone: +1-207-581-3930.

Please avoid calling except for truly urgent matters.

Email: sudarshan.chawathe@maine.edu

Use email only for messages unsuitable for the newsgroup. (See below.) Please use only this email address and put the string *ERS602* near the beginning of the Subject header of the message.

Web: <http://cs.umaine.edu/~chaw/>.

- Andrei V. Kurbatov

Office: Sawyer ERC.

Office hours: by appointment.

Phone: +1-207-581-2840

Email: akurbatov@maine.edu

Web: http://climatechange.umaine.edu/people/profile/andrei_kurbatov.

Online Resources

Class Web site: <http://cs.umaine.edu/~chaw/ers602/>

We will use the class Web site for posting assignments, readings, notes, and other material. Please monitor it.

Class Newsgroup: We will use the local USENET newsgroup `umaine.ers602` on the NNTP server `creak.um.maine.edu` for electronic discussions. If you are unfamiliar with USENET, you may find the Web interface at <http://creak.um.maine.edu/news/> useful as a quick way to get started. You may find further information on USENET at <http://en.wikipedia.org/wiki/Usenet>. The newsgroup is the primary forum for electronic announcements and discussions, so please monitor it regularly, and post messages there as well. Unless there is a reason for not sharing your question or comment, please *use the newsgroup, not email*, for questions and comments related to this course.

Class mailing list: *Please make sure you are on the class mailing list.* A sign-up sheet is circulated at the first class meeting. If you miss it, please contact me to get on the list. We will use this mailing list only for urgent messages because all other messages will go on the class newsgroup. I anticipate fewer than a dozen messages on this list over the semester.

Policies

Academic honesty (standard university wording): Academic dishonesty includes cheating, plagiarism and all forms of misrepresentation in academic work, and is unacceptable at The University of Maine. As stated in the University of Maine's online undergraduate Student Handbook, plagiarism (the submission of another's work without appropriate attribution) and cheating are violations of The University of Maine Student Conduct Code. An instructor who has probable cause or reason to believe a student has cheated may act upon such evidence, and should report the case to the supervising faculty member or the Department Chair for appropriate action.

Disabilities (standard university wording): If you have a disability for which you may be requesting an accommodation, please contact Ann Smith, Director of Disabilities Services, 121 East Annex, 581-2319, as early as possible in the term.

Special circumstances (standard university wording): In the event of an extended disruption of normal classroom activities, the format for this course may be modified to enable its completion within its programmed time frame. In that event, you will be provided an addendum to the syllabus that will supersede this version.

Readings

This list will be revised and annotated as the semester progresses. (In particular, the Assigned section below will grow.)

(Being updated; refer to the class newsgroup in the meanwhile.)

Assigned

1. John R. Taylor. *An Introduction to Error Analysis*. University Science Books, Sausalito, California, 2nd edition, 1982.
2. Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 3rd edition, 2011.

Assignments

Material will appear here as we move along the semester.

- Homework assignments:
 - Homework 1: described in class.
 - Homework 2: `hwq/hw02.pdf`.

Submission Instructions

All electronic submissions must be made using the upload interface at <http://cs.umaine.edu/~chaw/u/>. *Electronic submissions in all other forms, such as email or physical media, will be discarded and receive no credit.*

Uploaded *files must be named* following this template:

```
ers602-pr01-Lastname-Firstname-N.jar
```

The substrings `pr01` and `jar` are replaced by others depending on the material being submitted and N is an arbitrary 4-digit number, such as 4231. Multiple submissions, within reason, may be made by selecting different values of N .

If your upload is successful, you will be presented with a confirmation Web page similar to the following sample. You should record the reported MD5 checksum and timestamp.

```
SUCCESS: Please note the following for your records.
```

```
Successfully saved ers602-pr01-Aardvark-Alice-1389.jar.  
MD5 checksum: 09ee098b83d94c7c046d6b55ebe84ae1  
Timestamp: 2016-09-13 13:32:34
```

If you do not see something very similar then your submission is unsuccessful.

If (and only if) there are unexpected problems and you are unable to submit your work as above, then you should save your file on your own computer (with some backups), compute its MD5 checksum using the `md5sum` utility on Unix-like systems (or other similar tools), and submit the file name, time stamp, and MD5 checksum (only, not the file itself) by email with a suitable Subject header.