

CSC314: Data Mining

Dr. Robert Lowe

Spring, 2020

E-mail: robert.lowe@maryvillecollege.edu

Office Phone: 865-981-8169

Office Hours: MWF 1:00PM – 2:00PM, TR 3:00PM – 4:00PM Class Hours: TR 11:00 – 12:15

Office: SSC 214

Class Room: SSC 204

Course Description

Data mining is concerned with the extraction of information from large amounts of data. This project-based course introduces the concepts, issues, tasks and techniques of data mining. Topics include data preparation and feature selection, classification, clustering, evaluation and validation, and data mining applications.

Required Materials

- *Data Mining and Analysis – Fundamental Concepts and Algorithms* by Mohammed J. Zaki and Wagner Meira, Jr.
<http://www.dataminingbook.info>
- A shell account on cs.maryvillecollege.edu (alternatively, you could install R-Studio locally on your own computer).

Prerequisites

CSC313 and MTH321

Course Goals

- Learn and overcome the problems in working with real world data.
- Learn the difference between good and bad data mining practices.
- Explore feature extraction and dimension reduction.
- Use statistical models and machine learning algorithms to interpret data.
- Learn the R programming language.

Course Structure

Methods of Instruction

- Lecture
- Homework
- Projects

Grading

This course is graded using a weighted average among four categories. The assignments within each category are equally weighted and are all graded out of 100 points. Hence your final numeric grade is computed by finding the average of each category, and then multiplying them by the corresponding weight. The weights for each category are as follows:

Category	Weight
Homework	40%
Projects	60%

Letter grades will be assigned according to the following scale:

A+	96.7–100%	B+	86.7–90%	C+	76.7–80%	D+	66.7–70%	F	less than 60 %
A	93.3–96.7%	B	83.3–86.7%	C	73.3–76.7%	D	63.3–66.7%		
A-	90–93.3%	B-	80–83.3%	C-	70–73.3%	D-	60–63.3%		

Assessments

The standards of assessment in each grading category will be as follows.

Homework (40% of the final grade)

Homework will be assigned at various points throughout the semester. Most of the homework assignments will involve producing documents using R notebooks. You will turn in printouts of these notebooks.

Projects (60% of the final grade)

There will be three course projects. These will be completed in groups of 3-4 students, and each will result in a written report and a presentation. Your group will receive a collective grade on the projects. Each project is weighted equally.

Note that while your group may elect to have a single presenter on presentation days, the entire group must be present for the presentation. Anyone who does not attend on a presentation day will receive a reduced grade on the project.

Schedule

This is the tentative schedule for our course. There may be some slight modifications to the following according to the needs of the semester. Presentation days are boxed in on the calendars below.

Su	Mo	Tu	We	Th	Fr	Sa
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

- **Thu January 9** - Introduction to Data Mining
 - Begin Reading Chapter 1
- **Tue January 14** - Data Mining and Analysis
- **Thu January 16** - Introduction to R
- **Tue January 21** - Numeric Attributes
 - Read Chapter 2
- **Thu January 23** - Numeric Attributes
 - Homework 1 Assigned (Due January 30)
- **Tue January 28** - Categorical Attributes
 - Read Chapter 3
- **Thu January 30** - Categorical Attributes
 - Homework 2 Assigned (Due February 6)

February 2020

Su	Mo	Tu	We	Th	Fr	Sa
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29

- **Tue February 4** - Dimensionality Reduction
 - Read Chapter 7
- **Thu February 6** - Dimensionality Reduction

- **Tue February 11** - Eigenpets
 - **Thu February 13** - PCA and Project 1 (Due March 10)
 - Project 1 is Assigned (Due)
 - **Tue February 18** - Itemset Mining
 - Read Chapter 8
 - **Thu February 20** - Itemset Mining
 - Homework 3 is Assigned (Due February 27)
 - **Tue February 25** - Sequence Mining
 - Read Chapter 10
 - **Thu February 27** - Sequence Mining
 - Homework 4 is Assigned (Due March 5)
-

March 2020

Su	Mo	Tu	We	Th	Fr	Sa
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

- **Tue March 3** - Pattern and Rule Assessment
 - Read Chapter 12
- **Thu March 5** - Pattern and Rule Assessment
 - Homework 5 is Assigned (Due March 12)
- **Tue March 10** - Project 1 Presentations
- **Thu March 12** - Beginning Project 2 (Due April 2)
- **Tue March 17** Spring Break
- **Thu March 19** Spring Break
- **Tue March 24** - Representative-based Clustering
 - Read Chapter 13
- **Thu March 26** - Representative-based Clustering

– Homework 6 is Assigned (Due April 2)

- **Tue March 31** - Hierarchical Clustering

- Read Chapter 14

April 2020

Su	Mo	Tu	We	Th	Fr	Sa
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

- **Thu April 2** - Project 2 Presentations

- **Tue April 7** - Begin Project 3 and Clustering Validations

- Read Chapter 17

- Project 3 Assigned (Due April 24)

- **Thu April 9** - Clustering Validations

- **Tue April 14** - Project 2 Presentations

- **Tue April 21** - Support Vector Machines

- Read Chapter 21

- **Fri April 24 3:30PM** - Project 3 Presentations

Course Policies

Late Policy

No late work will be accepted under any circumstances (except as mercy and decency may dictate in extremely rare events).

Extra Credit

No extra credit will be given under any circumstances.

Excused Absences

In some cases, absences may be excused. These include:

- School Sanctioned Events (Sports, Concerts, etc.)
- Severe Illness
- Family Emergencies
- Court Appearance / Jury Duty

In the case of a school event, notice must be given at least one week prior to the absence. The notice must include a signed note from the faculty or staff member in charge of the event. This note must be given in physical form, electronic notes will not be accepted. In the case of illness, a doctor's note is required. Note that except in extreme circumstances, doctor's appointments do not qualify as a valid reason to miss a class. Please be respectful of the other students, and schedule appointments during your free time.

Family emergencies will require some form of proof. Where possible, you must give advance notice of missing a class. The exception to this would need to be fairly severe, and hopefully it will not come up. For court appearances and/or jury duty, you must provide a copy of your summons. You may redact any details you wish, save for the actual date and time of your appearance. Court appearances must be cleared at least one week in advance.

Making Up Excused Absences

Should you be in a situation in which you receive an excused absence, this in no way will extend your due dates (excepting extreme emergencies). You must make up any test at a designated time prior to your excused absence. Also, homework or projects must be submitted prior to the class period in which they are due.

Communication and Extra Help

You are always welcome at office hours for help with any questions you may have about the course. For help at other times during the day, stop by or call my office to see if I'm available. You can also contact me by email, but often I can better help you face to face and may respond with a request that you come to see me. Note that I do not typically respond to email between 5 p.m. and 8 a.m. You may make appointments to see me at other times if your schedule does not permit you to attend my office hours.

Plagiarism and Cheating

You are expected to do your own work. Never submit the work of others, never give unauthorized assistance to others, do not use unauthorized aids during exams, and do not ask for help from other faculty members without the approval of your professor. Plagiarism and cheating are serious offenses that will not be tolerated. Explanations regarding these offenses and how they are handled can be found in the MC Student Handbook at

<https://www.maryvillecollege.edu/academics/catalog/handbook/section-nine/>.

You are expected to read and understand these policies. Offenses on specific assignments, quizzes, or exams will result in a score of 0 on the relevant assignment, and a letter of censure will be placed in your college file. Repeat offenses will result in further disciplinary action, including the possibility of failing the course.

Students with Disabilities

Any student who feels s/he may need learning or physical accommodation(s) based on the impact of a disability should contact Services for Students with Disabilities to discuss your specific needs. Please contact 981-8124 to coordinate reasonable accommodations for students with documented disabilities. The Disability Services office is located in the Learning Center in the basement of Thaw Hall. Undocumented disabilities will not be accommodated.