

University of Wisconsin - Stevens Point
 College of Letters and Science
 Department of Computing and New Media Technologies
 Spring 2019 – Version 1.3

Course: Data Mining (DAC 310) #41586
 Books: Data Mining Techniques, 3rd Edition, Linoff & Berry
 Class Time: Tu Th 2:00-3:50 pm (Room Science B238)
 Professor: Dr. Kurt A. Pflughoeft (Floog'heft)
 Office: Science B231
 Office hours: M, Tu, Th 1:00-2:00 pm, W 3:30-4:30 pm by appointment
 Contact: kpflugho@uwsp.edu

Course Description: Organizations and business are overwhelmed by the flood of data continuously collected into their data warehouses and arriving from external sources – the Web above all. Traditional exploratory techniques may fail to make sense of the data, due to its inherent complexity and size. Data mining and knowledge discovery techniques emerged as an alternative approach, aimed at revealing patterns, rules and models hidden in the data, and at supporting the analytical user to develop descriptive and predictive models for a number of business problems. This course focusses on the main applications scenarios of data mining to challenging problems in the broad CRM domain - Customer Relationship Management.

<u>Week #</u>	<u>Dates</u>	<u>Topic*</u>	<u>Assignments</u>
1.	Jan 22,24	Introduction to DM & Knime	Chapters 1&2, Lab 1
2.	Jan 29,31	DM Process & Stats Part 1	Chapter 3, Lab 2, HW 1
3.	Feb 05,07	Stats Part 2	Chapter 4, Lab 3
4.	Feb 12,14	Profiling and Prediction	Chapter 5, Lab 4 HW 2
5.	Feb 19,21	DM via Statistics – Part 1	Chapter 6, Lab 5
6.	Feb 26,28	DM via Statistics – Part 2	Lab 6, HW 3
7.	Mar 05,07	Decision Trees	Chapter 7, Lab 7
8.	Mar 12,14	Review & Midterm	
	Mar 19,21	Spring Break	
9.	Mar 26,28	DT & ANN	Chapter 8, Lab 8, HW 4
10.	Apr 02,04	ANN & Nearest Neighbor	Chapter 9, Lab 9
11.	Apr 09,11	Clustering	Chapters 12 & 13, Lab 10, HW 5
12.	Apr 16,18	Market Basket Analysis	Chapter 15, Lab 11
13.	Apr 23,25	Link Analysis	Chapter 16, Lab 12, HW 6
14.	Apr 30,May 2	Text Mining	Chapter 21, Lab 13
15.	May 07,09	Derived Variables & Review	Chapter 19
16.	May 14	Final Exam 12:30-2:30 pm	

Schedule Footnotes:

- The class meetings will be a mixture of lecture and lab activities with a break in between.
- Typically, I like to keep the book material separate from lecture as another viewpoint. However, due to the book's length (845 pages), I will incorporate some of that content into lecture.
- This schedule is a guide to the coverage of topics. The instructor reserves the right to alter the presentation schedule as necessary to benefit the class.

Check D2L for updates to syllabus and due dates for homeworks and quizzes.

Course Outcomes - Given a successful conclusion of this course, students will be able to:

- 1) Have a broad understanding of the principles and the concepts of data mining methods and their applications.
- 2) Ability to apply creative thinking to resolve complex problems or issues as well as summarizing complex multivariate data and creating visual summaries of such data
- 3) Explain the link between descriptive and predictive data mining to support good decision making
- 4) Examine and compare the differences between several supervised techniques for decision makers by explaining results in either a technical or non-technical vernacular.
- 5) Analyze data sets by applying classification and cluster analysis methods and use their results to create an action plan for the management
- 6) Apply market basket analysis to the sales data of a company, synthesize the results for a professional data mining report
- 7) Demonstrated level of knowledge and technical expertise in data mining activities, including cleaning and transformation of data; presentation of results of mining and modelling to possible users
- 8) High-level research, analytical and conceptual skills and ability to apply these skills in development of models and client profiling
- 9) Apply the concepts introduced in this course to data sets using KNIME & R

Distribution of Points

Midterm	:20%
Final	:25%
Labs	:20%
Quizzes	:10%
Homeworks	:20 %
Attendance	:05%

Knime Analytics Platform is the leading open solution for data-driven innovation, helping you discover the potential hidden in your data, mine for fresh insights, or predict new futures. The enterprise-grade, open source platform is fast to deploy, easy to scale and intuitive to learn. With more than 1000 modules, hundreds of ready-to-run examples, a comprehensive range of integrated tools, and the widest choice of advanced algorithms available, KNIME Analytics Platform is the perfect toolbox for any data scientist.

R: The open source R package provides a complete analytical environment and is justifiably viewed as the de facto language in the data science community. It is in fact the fastest growing language on the StackOverflow developers site and is currently ranked in 5th place in IEEE language rankings.

POLICIES

Academic Standards - UW-Stevens Point values a safe, honest, respectful, and inviting learning environment. In order to ensure that each student has the opportunity to succeed, we have developed a set of expectations for all students and instructors. This set of expectations is known as the Community Rights and Responsibilities document, and it is intended to help establish a positive living and learning environment at UWSP. Click here for more information: <http://www.uwsp.edu/dos/Pages/AcademicMisconduct.aspx> Academic integrity is central to the mission of higher education in general and UWSP in particular. Academic dishonesty (cheating, plagiarism, etc.) is taken very seriously. Don't do it! The minimum penalty for a violation of academic integrity is a failure (zero) for the assignment. For more information, see the "Student Academic Standards and

Disciplinary Procedures” section of the Community Rights and Responsibilities document, UWSP Chapter 14. This can be accessed at: <http://www.uwsp.edu/dos/Documents/CommunityRights.pdf> - page=11

ADA Statement - The Americans with Disabilities Act (ADA) is a federal law requiring educational institutions to provide reasonable accommodations for students with disabilities. For more information about UWSP’s policies, check here: <http://www.uwsp.edu/disability/Pages/faculty/lawAndPolicy.aspx>. If you have a disability and require classroom and/or exam accommodations, please register with the Disability and Assistive Technology Center at the beginning of the course and then contact me. I am happy to help in any way that I can. For more information, please visit the Disability and Assistive Technology Center, located on the 6th floor of the Learning Resource Center (the Library). You can also find more information here: <http://www.uwsp.edu/disability/Pages/default.aspx>.

Attendance Policy - Attendance will be taken randomly in lecture/lab and will count towards your grade! I rarely lecture “STRAIGHT FROM” the book.

Audio/Visual Recording Policy - Electronic recording of lectures (taping) is prohibited unless receiving prior written approval from the instructor. Approval will be granted only for self-study purposes. You are allowed to take pictures of whiteboards, blackboards or screens of my lecture material, if need be.

Average Time Investment/Workload Policy Statement

DAC 310 meets twice a week; each meeting is 110 minutes or about 4 hours per week or 64 hours per semester. Additionally, you should expect to spend up to 8 hours per week, on average, on outside class work including homework, quizzes and chapter reading assignments.

Classroom conduct – Please mute cell phones and any audible device during classes. Please do not hold private conversations or text while I am lecturing as it is a distraction. No FOOD or DRINKS are allowed in the lab.

D2L –

- Recorded grades as well as lecture materials (syllabus, PowerPoint class outlines, etc.) will be available on our D2L course site.
- It is your responsibility to check that your grades are posted correctly on D2L. Questions about any posted grade must be raised within TWO weeks of posting. Beyond this time frame, all grade postings are considered correct and final. The D2L site is not available after the final exam.
- It’s best to download attachments rather than view directly in the browser.
- USE the OneDrive to temporarily save your files – if need be.
- **News feed** on D2L is the main communication tool (not email!) - consider the text msg option
- The UW system is in the process of moving to a different LMS called Canvas

Drop Policy - In accordance with the rules stated by the College of Letters and Science. I will **NOT** personally drop a student - you are responsible for filling out all the forms.

Email Policy

- I try to answer questions in a timely manner but if you haven't received a response from me by the end of the next business day, please resend the email.
- If your email is only informative in nature, such as you are missing a class, I usually don't reply to those emails but rather just file them. If your email has a question or issue that needs to be addressed, I will reply to it. Likewise, if I send you an email that requires a response, please do so.
- Please include "DAC 310" as part of your subject line.

Exam Policy - Except for documented emergencies, no late or makeup in-class exercises, exams and quizzes will be given.

Grade Policy - The following scale can always be used to estimate your grade

Percentage breakdown for semester grades (weighted point totals)

A = 93-100%	B- = 80-82.99%	D+ = 67-69.99%
A- = 90-92.99%	C+ = 77-79.99%	D = 63-66.99%
B+ = 87-89.99%	C = 73-76.99%	D- = 60-62.99%
B = 83-86.99%	C- = 70-72.99%	F = < 60%

*Instructor reserves the right to implement a curve which is beneficial to the students.

Homework – due dates available in D2L. Late homeworks are discounted 20 percentage points per day.

Labs –are scheduled the second hour of class and usually have in-class exercises. For lab assignments, you should turn in a Word document which lists your code, and one or more screen shots (or relevant copy/pastes) of the program's output to demonstrate the program works correctly. Always check the lab/homework instructions as other materials may also need to be turned in. If you have extra lab time, you are encouraged to work on your DM assignments. Number of lab assignments may vary from schedule. Lab assignments must be started during class.

Lecture Notes – electronic version of the notes is available for some topics, however, I strongly encourage you to take good notes as that has been shown to reinforce memory recall.

News – Always check the news item on D2L to find the latest announcements concerning the class.

Software – Lab Virtual Desktop or install RStudio and Knime on your PC. I can help you with the Knime and R install on your laptop. 8 gig of RAM should be sufficient for an academic environment but data analysts in industry often require more as all data must be loaded in primary storage for R.

For laptop installs:

Use UWSP Software Center OR

Download R at <https://cran.r-project.org/>

Download RStudio: <https://www.rstudio.com/products/rstudio/download/#download>

Download Knime: <https://www.knime.com/downloads>

Plagiarism Policy - All assignments and tests should represent YOUR work otherwise you will not receive any credit for that portion of your grade. Disciplinary actions will be pursued for serious offenses – see Academic Standards.

Quiz Policy – quizzes are meant to test your understanding about topics that were currently presented. Quizzes will be take-home but you are NOT allowed to collaborate with others. You may use other resources such as google. For open-ended questions, be careful not to plagiarize. There will be 6 quizzes check D2L for due dates. No late quizzes are accepted. I do reserve the right to hold in-class quizzes if the students are not doing the required work outside of the classroom.

Texting/emailing – you are not allowed to email or text during the lecture component of class unless it is an emergency. Students who need to text may do so outside the classroom. Many studies have pointed that texting is a bit like an addiction, causes a lack of focus and is strongly correlated with poor grades.

University Emergency Preparedness –

In the event of a medical emergency call 9-1-1 or use Red Emergency Phones. Offer assistance if trained and willing to do so. Guide emergency responders to victims.

In the event of a tornado warning, proceed to the lowest level interior room without window exposure. See www.uwsp.edu/rmgt/Pages/em/procedures/other/floor-plans.aspx for floor plans showing severe weather shelters on campus. Avoid widespan structures (gyms, pools or large classrooms.)

In the event of a fire alarm, evacuate the building in a calm manner. Stay 200 yards away from the building. Notify instructor or emergency command personnel of any missing individuals.

Active Shooter – RUN/ESCAPE, hide, fight. If trapped hide, lock doors, turn off lights, spread out and remain quiet. Call 9-1-1 when it is safe to do so. Follow the instructions of emergency responders.

See UW-Stevens Point Emergency plan at <https://www.uwsp.edu/rmgt>

