COMP SCI 5402 Introduction to Data Mining

Summer 2020 Syllabus

Version 1.0

Instructor: Perry B. Koob
eMail: koobp@mst.edu

Course time: Monday, Tuesday, Wednesday, Thursday, and Friday from 12:00pm to 1:00pm.
Course room: Online-delivery
Live and recorded lecture can be accessed at:
https://mst.mediaspace.kaltura.com/
Help Sessions: Zoom by appointment

Course Outcomes

The key objectives of this course are two-fold:

1. to teach the fundamental concepts of data mining and
2. to provide extensive hands-on experience in applying the concepts to real-world applications.

The core topics to be covered in this course include classification, clustering, association analysis, data preprocessing, and outlier/novelty detection.

Intended Course Audience/Prerequisites

The intended audience for this course is students who have a grade of "C" or better in all of Comp Sci 2300, Comp Sci 2500, and one of Stat 3113, Stat 3115, Stat 3117 or Stat 5643.
Course Materials

Course Text:
Data Mining: Practical Machine Learning Tools and Techniques
(Morgan Kaufmann Series in Data Management Systems)
4th Edition
(Witten, Frank, Hall, and Pal – Morgan Kaufmann)

ISBN-10: 0128042915

Available Online at Safari Books
https://proquest.safaribooksonline.com/9780128043578

Papers:
A selection of papers will be assigned for reading. These papers are chosen to provide a deeper understanding of the topics, and allow students to become familiar with using papers as a resource to learn new data mining techniques.

Community articles:
R and Python have rich user communities that have numerous member written articles and tutorials that will be referenced.

Canvas:
Canvas, the Learning Management System (LMS) will be used to provide information including administration of homework and grades. This can be accessed through http://canvas.mst.edu and additional help can be sought from the IT Helpdesk (http://it.mst.edu/help-desk/).

Kaltura MediaSpace:
If you are a distance student and would like to watch the class live online, you can use Kaltura MediaSpace. The live video can be accessed at https://mst.mediaspace.kaltura.com/.

Panopto:
Panopto is an online video platform for businesses & universities. If you are a distance student, this will be your main way of receiving course content. Panopto may also be used for pre-recorded content for all students. This can be accessed through https://mst.hosted.panopto.com/ and through links in canvas as well.
Tentative Outline

1 Week What is Data Mining (Ch 1)

   Data Mining Ethics and Bias

   The vocabulary, inputs, and outputs of Data Mining (Ch 2 and 3)

   Data Mining Tools

1 Week Evaluation Techniques for Data Mining (Ch 5)

2 Weeks Basics of Classification, Numeric Prediction, Instance Based Learning, and Clustering (Ch 4)

1 Week Data Transformation Techniques (Ch 8)

3 Weeks Data Mining Methods (4,6,7,9,10)

Grading

Discussion and Participation 10%
Homework Programs and Assignments/or Quizzes 20%
Semester Project 20%
Exam 1 25%
Exam 2 25%

Maximum breaks:
   90% or more = A
   80% points or more = B
   70% points or more = C
   less than 70% = F

These values represent the maximum required to get the associated letter grade. I reserve the right to lower individual breaks.

Discussion and Participation
Discussions are a means to document your learning process. Your participation is essential to your success in the class. You are expected to contribute to the Canvas Discussion board and to in-class discussions. The discussion board topics are a way for you to make the course what you want. For these discussions you may be asked to briefly discuss the reading, a topic
discussed in class, or quickly summarize a paper related to class that you have been asked to pick out. Each topic will help you gain a deeper understanding of the applications and challenges of data mining.

Graded discussions will be posted on Canvas and clearly marked as such. You will be required to post a reply on these graded discussions or you will receive no discussion points. The replies must be brief, yet thoughtful answers to the topics posted. I expect a reply to be around a paragraph or at least three sentences. After you have posted a reply to the discussion, you will be expected to read the replies of other classmates and post a meaningful response to the ones you either find the most interesting, or ones you feel you can contribute additional information to.

Each Canvas Discussion will provide detailed information on what the post is looking for along with a rubrik.

**Exams**
There will be two exams for the course. The exams will test your understanding of the basic concepts covered in the course and will be a selection of true/false questions, multiple choice questions, and/or short essay-type questions. There will be no comprehensive final exam at the end of the semester. However the topics in this class do build upon each other.

A week prior to the exam a practice exam will be made available. The practice exam will be made using questions that will be used to create the exam.

**Homework Programs and Assignments/or Quizzes**
Homeworks Programs and Assignments are designed to allow you to apply the theory you have read in the course material or that has been presented in class.

**Semester Project**
The semester project is designed to allow students to apply the theory from the class on a larger scale than the Homework Programs. Additionally the project will give students real world experience of the ups and downs of working collaboratively on a data mining team. It will also give them experience communicating complex concepts and ideas with others.
Course Requirements/Policies

Late Work
Homework can be submitted after the due date for partial credit, as long as they are submitted before the assignment has been returned, for a deduction of 25% if the homework is submitted within 24 hours past the due date. The homework will accrue an additional 25% deduction for each additional 24hrs until a total of 100% deduction has been accumulated, or until such time as the answers to the homework has been released. This is not to say that the answers to the homework will not be released until 4 days after the due date. The answers may be released sooner or later. All deadlines will be entered within Canvas for the given assignments.

Attendance
Your learning of the material requires a concerted effort on your part. Attendance at, and participation in, class show your commitment to learning. If your attendance and grades appear significantly lacking, you may be administratively dropped. Conversely, if you make a strong effort toward attending and participating in class, that may be considered in assessment of final grades.

Respect for Other
You will be asked to post your opinions of observations of various topics. I love the diverse set of opinions and approaches we have on the discussion posts and feel it brings a lot of value to the class. However, you must be respectful of others when you post your comments. This does not mean censorship, this means compassion, the greatest quality we can show people who do not agree with us.

I don't want anyone to feel like they are being singled out because of their gender, ethnicity, religion, culture, beliefs, gender representation, sexual orientation, home, family, occupation, military service, age, height, weight, upbringing, socioeconomic status, life experiences, diet choice, pets, choice of entertainment, or any other reason I am not thinking of right now. It is a serious matter that can quickly escalate to disciplinary actions.

So, please consider how your words could be taken by others.

Cheating
You may work together with classmates on homework as long as you turn in your own work. You are not allowed to work with others on discussion posts or exams in person or digitally. If it has been discovered that you have collaborated with others on your discussion post or the exams, all of you will receive a zero and the Academic Dishonesty Policy will apply.
Campus Policies

Disability Support Services:
Any student inquiring about academic accommodations because of a disability should contact Disability Support Services (http://dss.mst.edu) so that appropriate and reasonable accommodative services can be determined and recommended. Disability Support Services is located in 204 Norwood Hall. Their phone number is 341-4211 and their email is dss@mst.edu.

If you have a documented disability and anticipate needing accommodations in this course, you are strongly encouraged to meet with me early in the semester. You will need to request that the Disability Services staff send a letter to me verifying your disability and specifying the accommodation you will need before I can arrange your accommodation. Accommodation arrangements may require time to set up or coordinate, so if you have accommodations it is best to notify me early about them even if you do not intend to use them.

Academic Dishonesty:
(http://registrar.mst.edu/academicregs/index.html) Page 30 of the Student Academic Regulations handbook describes the student standard of conduct relative to the System's Collected Rules and Regulations section 200.010, and offers descriptions of academic dishonesty including cheating, plagiarism or sabotage. Additional guidance for faculty, including a description of the process for dealing with issues related to academic dishonesty, is available on-line at http://ugs.mst.edu. In particular, you must work alone on all tests, exams, and individual programming assignments unless otherwise stated.

If you have any questions about the Campus Policies information listed above, please contact the Office of Undergraduate Studies at 573-341-7276.

Title IX:
Missouri University of Science and Technology is committed to the safety and well-being of all members of its community. US Federal Law Title IX states that no member of the university community shall, on the basis of sex, be excluded from participation in, or be denied benefits of, or be subjected to discrimination under any education program or activity. Furthermore, in accordance with Title IX guidelines from the US Office of Civil Rights, Missouri S&T requires that all faculty and staff members report, to the Missouri S&T Title IX Coordinator, any notice of sexual harassment, abuse, and/or violence (including personal relational abuse, relational/domestic violence, and stalking) disclosed through communication including but not limited to direct conversation, email, social media, classroom papers and homework exercises. Missouri S&T’s Title IX Coordinator is Vice Chancellor Shenethia Manuel. Contact her directly (manuels@mst.edu; (573) 341-4920; 113 Centennial Hall) to report Title IX violations. To learn more about Title IX resources and reporting options (confidential and non-confidential) available to Missouri S&T students, staff, and faculty, please visit http://titleix.mst.edu.