



Marshall University Syllabus
College of Engineering and Computer Sciences
Computer Science

Course

CS 415 Data Mining cross-listed with CS515

Course Description

Covers (1) the process of knowledge discovery, (2) algorithms (association rules, classification, and clustering), and (3) real-world applications. Focuses on efficient data mining algorithms and scaling up data mining methods.

Credits

3

Prerequisites

CS215 and CS410

Term/Year

Fall 2024

Class Meeting Days/Times

Tuesday and Thursday, 2:00pm–3:15pm

Location

WAEC 3123

Academic Calendar

For beginning, ending, and add/drop dates, see the [Marshall University Academic Calendar](https://www.marshall.edu/academic-calendar/) (URL: <https://www.marshall.edu/academic-calendar/>).

Instructor

Ananya Jana

Contact Information

- Office: WAEC 3103
- Office Hours: By appointment any day, and 10:45am–2:00pm Tuesday and Thursday. Wednesday 8:30am - 12:00pm
- Office Phone: [Enter office phone number]

- Marshall Email: jana@marshall.edu

Health and Safety Information

All members of the Marshall University community are expected to always observe health and safety protocols. This includes general health and safety protocols as well as specific protocols that might emerge in response to community and campus health conditions.

Campus Carry Policy

University Policy, UPGA-12 (Campus Carry Policy) derives its authority from West Virginia State law, including the Campus Self-defense Act (W. Va. Code § 18B-4-5b). It pertains to the exercise of Concealed Carry on Marshall University's campus, except in designated areas, by individuals with a valid permit to Conceal Carry.

Individuals who choose to Conceal Carry are responsible for knowing and understanding all applicable federal, state, and local laws and Marshall University Board of Governors Rules, University Policies, and Administrative Procedures. University Policy, UPGA-12 applies to areas of campus and buildings that are directly under the possession or control of Marshall University.

Concealed Handguns are not observable to others and must be holstered and concealed on the body of the permit holder or in a personal carrier, such as a backpack, purse, or other bag that remains under the exclusive and uninterrupted control of the permit holder. This includes wearing the personal carrier with a strap, carrying or holding the personal carrier, or setting the personal carrier next to or within your immediate reach at all times. If your participation in class activities impedes your ability to maintain constant control of your Handgun, please make alternate arrangements prior to coming to class.

Faculty Office

NOTICE: University Policy, UPGA-12 (Campus Carry Policy) defines Sole Occupancy Offices as areas that may restrict Concealed Carry. Please be aware that my office is a Sole Occupancy Office and this statement serves as notice that concealed weapons or handguns are not permitted in my office. If you plan to attend a meeting in my office or to drop by my office, secure your weapon or handgun appropriately before you arrive.

Required and/or Recommended Texts and Materials

Required Texts and Materials

Introduction to Data Mining By Pang-Ning Tan, Pearson International, ISBN 0-321-42052-7

- Lecture Slides

- Research Paper available at My MUOnline.

Course Student Learning Outcomes

The table below shows the following relationships: How each student learning outcome will be practiced and assessed in the course.

Course student learning outcomes	How students will practice each outcome in this course	How student achievement of each outcome will be assessed in this course
Students will be able to understand data mining methodology and techniques (a,b) ¹	1. In class discussion. 2. Presentations 3. Research Articles	1. Popup Quizzes 2. Graded Exam
Students will apply various procedure and methods for data classification (b,e,i)	1. Presentations 2. Research Articles 3. Data Mining Tools 4. Literature 1. Graded Project 2. Graded Exam	Exams 1. Graded Project 2. Graded Exam
Students will be able to utilize a state-of-the-art commercial data mining package (a,j,k)	1. Open Source Data mining tools/Packages	1. Graded Project
Students will be able to Interpret, verify and validate the results of data mining techniques (b,e)	2. Open Source Data mining tools/Packages 3. Simulator	1. Graded Project
Students will survey applications of data mining (g,h)	1. Literature 2. In Class Discussion	1. Graded Project 2. Graded Assignment

Course Requirements/Due Dates

- Course announcements will be posted on Blackboard. Email communication will be sent to your official Marshall email address. Please check these sources regularly to avoid missing any news.
- You are required to submit assignments and in-class exercises individually unless otherwise noted.
- You are required to take the exams. Make-up exams will only be given when there are documented circumstances that prevented you from

taking the exam at its scheduled day and time. Tentative exam dates are shown below; the final confirmation of each exam date will be at least one week in advance.

Grading Policy

The components of the course will be weighted as follows.

- Attendance 5%
- Pop-up Quizzes 15%
- Three Assignments 20%
- Midterm 20%
- Paper Presentation 10%
- Paper Critique 5%
- Project Proposal 5%
- Mid-Project Progress 5%
- Final Project Paper 15%

It is expected for students to attend all classes. Only University-excused absences will be accepted.

Attendance accounts for 5% of the overall course grade.

-1 Marks for the first absence.

-2 Marks for the second absence.

-3 for the third absence

-4 for the second absence.

Note: Accumulate -(ve) score will be added to students overall grade.

The overall course letter grade will be assigned using the following system.

- A: 100%-90%
- B: 89% -80%
- C: 79%-70%
- D: 69%-60%
- F: 59%-0%

Academic Honesty Policy

All work that is submitted in this course must represent your own understanding and effort. The very act of writing your solutions to homework is an important part of the learning process. The main purpose of the homework is to help you build an internal conception and understanding of the course content.

Collaboration on the conceptual parts of the homework is strongly encouraged, but direct copying or close paraphrasing of others' work is not allowed. Discuss the concepts with others, then write your own answers based on your own understanding.

Collaboration on programming assignments should be limited to the conceptual aspects. Do not show your code to other students or copy the code of other students. Do not directly copy code from the internet or anywhere else.

Exams

This class will have three assignments and one mid-term exam. The exam is given during class. You may not use any external resources or communicate with others during the exam. The course will consist of a semester-long project. The students enrolled in CS 515 will work on additional assignments such as reading research paper, summarizing research paper and writing research work critique.

Assignments

All assignments are due by the Saturday 11:59pm in that particular week. Assignment may involve writing code using scripting language such as Python. Late submission is not allowed.

Tools/Softwares

Anaconda, Jupyter notebooks, Spyder IDE

Use of external resources on homework

You are welcome to use external resources to help with the homework, such as books, chat boards, etc. However, if you directly use content from any source other than the course notes and textbooks, you must include a citation. The exact format of the citation is not very important as long as it provides enough information for me to easily locate the source you used. This includes a working link if the information is online.

Keep in mind that the goal of homework is to help you build an internal understanding of the content. That understanding will be necessary to succeed in the course and to apply course material in other contexts. Homework is a means to an end, not an end in itself.

Use of AI on homework

You are allowed to use Generative AI or other resources on the homework assignments with an appropriate citation, as with other external resources.

- Generative AI can “hallucinate” (produce false information). You are responsible for ensuring the accuracy of any AI-generated content.
- It is your responsibility to ensure that answers you submit are coherent and address the question that was asked. Any computer code you submit must run without errors and meet the specifications of the problem.

Attendance/Participation Policy

Attendance at all class meetings is expected, with the same standards as a place of employment. If you cannot attend a class, you should contact me as soon as possible by email – preferably before the class meeting.

- Because all homework submissions will be on MU Online, you can submit assignments even if you are not in class, as long as you submit them by the due date and time.
- We will follow the University excused absence policy for the purposes of make up assignments. A link is provided under "University Policies". Please contact me as soon as possible if you would like to discuss a possible excused absence.
- Make up examinations, including the final examination, are available only for excused absences, and must be taken within two weeks of returning to class after the absence.

Generative Artificial Intelligence (AI) Policy for Use in the this Course

Moderate Use - Generative AI is permitted/encouraged in some ways, with proper attribution, but prohibited in other ways.

Students are allowed, and even encouraged, to use Generative AI in some ways but are prohibited from using it in other ways. Keep in mind that any content produced by generative AI can “hallucinate” (produce false information), so students are responsible for ensuring the accuracy of any AI-generated content. For information on citing AI, please see [MU Library's citation website](https://libguides.marshall.edu/plagiarism-AI/cite) (URL: <https://libguides.marshall.edu/plagiarism-AI/cite>). Students should not use generative AI in any way that would violate the [Student Code of Conduct](https://www.marshall.edu/student-conduct/) (URL: <https://www.marshall.edu/student-conduct/>).

Students are **permitted and encouraged** to use generative AI in the following

ways:

- **Brainstorming:** You may use generative AI to stimulate creativity, generate ideas, or brainstorm topics for papers, presentations, and discussions. The generated content must serve as a stepping stone, not a final product.
- **Citation Assistance:** AI tools can be used to manage, format, and organize citations and references, promoting adherence to academic writing standards and specific style guides required for individual assignments.
- **Grammar and Style Checking:** AI-powered writing enhancement tools may be used to help with spelling, grammar, syntax, and stylistic errors.
- **Concept Understanding:** Generative AI can be used to explain or simulate concepts taught in class, aiding in a deeper understanding.
- **Research Assistance:** AI can be used to conduct initial research, compile data, and summarize articles, books, or papers. It should not replace traditional research methods but rather enhance them.

You **may not** use generative AI in coursework in the following ways:

- **Plagiarism:** Using AI-generated content as your original work without attribution. This includes essays, papers, presentations, and exam answers.
- **Data Manipulation:** Using AI tools to alter data or create misleading information.
- **Misrepresentation of Skills:** Using generative AI to complete tasks that are meant to assess your knowledge and skills.
- **Confidentiality Breach:** Using AI tools that might violate university policies or laws related to data privacy and confidentiality.

Metacognitive Reflection. In addition to a proper citation, the student should include the following statement with any assignment where generative AI is used for assistance.

"I used generative AI platform [INSERT NAME OF PLATFORM, SUCH AS CHAT GPT] for assistance in the following ways on this assignment: [INSERT WAYS USED, such as brainstorming, citation assistance, grammar and style checking, concept understanding, and research assistance,

University Policies

By enrolling in this course, you agree to the University Policies. Please read the full text of each policy (listed below) by going to [MU Academic Affairs: University Policies](https://www.marshall.edu/academic-affairs/policies/). (URL: <https://www.marshall.edu/academic-affairs/policies/>)

- Academic Dishonesty Policy
- Academic Dismissal Policy
- Academic Forgiveness Policy
- Academic Probation and Suspension Policy
- Affirmative Action Policy

- Pre-Finals Week Policy
- D/F Repeat Rule
- Excused Absence Policy for Undergraduates
- Inclement Weather Policy
- Sexual Harassment Policy- Title IX prohibits the harassment of students based on sex, which includes pregnancy, childbirth, and related conditions. This includes that students will not be penalized for taking medically necessary leave related to pregnancy, childbirth, or related conditions. Marshall's Title IX Office may be contacted at TitleIX@marshall.edu
- Students with Disabilities (Policies and Procedures)
- University Computing Services Acceptable Use Policy

Course Schedule

The assignments highlighted in green i.e. the paper presentation, paper critique and paper summary are only for CS 515 in addition to the regular assignments, mid-term and project.

Week	Topics	Assignment Due
1: 8/20, 8/22	Introduction to Data Mining	
2: 8/27, 8/29	Understanding Data	
3: 9/3, 9/5	Basic Classification: Decision Trees	Assignment 1
4: 9/10, 9/12	Model Overfitting	
5: 9/17, 9/19	Ensemble Learning	
6: 9/24, 9/26	Artificial Neural Networks	Assignment 2
7: 10/1, 10/3	Imbalanced Classes	
8: 10/8,	K Nearest Neighbor Algorithm	Project Proposal (presentation) Paper Presentation Paper Critique
9: 10/15, 10/17	Naive Bayes Classification	Assignment 3
10: 10/22, 10/24	Clustering	

Week	Topics	Assignment Due
11: 10/29, 10/31	Support Vector Machines	Mid-Term
12: 11/5, 11/7	Association Analysis	
13: 11/12, 11/14	Anomaly Detection	Mid-project progress (presentation)
14: 11/19, 11/21	Fully Supervised, Semi-supervised and Self-supervised methods	
15: 12/3, 12/5	Neural Networks	
	Final Exam Tuesday 12/12, 12:45pm - 2:45pm	Final Project is due - Report - Code/repo - Webpage Paper Summary is due