WICHITA STATE UNIVERSITY

IME 734, Introduction to Data Mining and Analytics, Fall, 2023

- Instructor: Dr. Laila Cure
- Department: Industrial, Systems and Manufacturing Engineering
- Office Location: Engineering Building, 120
- Telephone: 3169787626
- Email: laila.cure at wichita dot edu
- Preferred Method of Contact: Email. Allow for 24 hours (M-F) for a response.
- Student Office Hours:
  - Mondays, **In person only**: 12:30 pm – 1:30 pm, In Person.
  - Wednesdays, **Virtual only**: 12:30 – 1:30 pm Click here to join the meeting
  - Other times by request.
- Classroom Day/Time: John Bardo Center 340, TR 2:00 pm – 3:15 pm
- Prerequisites: IME 254 Engineering Prob./Stat. I. or equivalent, Math 511 Linear Algebra or equivalent, Basic knowledge Internet search, e-mail, presentation, and Excel spreadsheets. The course requires students to self-learn a selected software package to apply the techniques learned during the course. Get started with the software packages used in examples:
  - Excel Pivot tables and charts:
    - https://www.excel-easy.com/data-analysis/pivot-tables.html
    - https://www.excel-easy.com/examples/pivot-chart.html
  - Tableau
    - Getting started: https://www.tableau.com/learn/training
    - LinkedIn Learning Video: “Exploratory analysis with Tableau”
  - RStudio (free open-source software):
    - Getting started: https://www.youtube.com/watch?v=lVKMsaWju8w
    - LinkedIn Learning Video: “Using Rstudio”
  - R Tutorial: https://www.statmethods.net/r-tutorial/index.html
- Teaching Assistant: None
- TA Contact: N/A

How to use this syllabus

This syllabus provides you with information specific to this course, and it also provides information about important university policies. This document should be viewed as a course overview; it is not a contract and is subject to change as the semester evolves. Changes to the syllabus will be announced through Blackboard and the updated document will be posted in the syllabus folder.

Student Office Hours

My student office hours are listed on page one of the course syllabus. Please utilize this time to meet with me for course related discussions: asking for extra help with lab assignments, seeking further clarification of material presented in class, and follow up on aspects of the class you are interested in. These hours are not required for the course but are available to help support you on your path to success.

Course Description

Graduate level course introducing the theory, application, and interpretation of basic methods for analyzing existing datasets. Topics include describing datasets, data preprocessing, data exploration, visualization, and basic methods for regression, classification, cluster, and association analysis. The course will focus on learning the data mining tasks that each method addresses, the assumptions of each method, the inputs needed, the outputs, interpretation of results, and evaluation of the quality of the analysis. The course includes a flipped-lab component based on student-selected datasets. Throughout the course, students will be introduced to Tableau and RStudio to implement the methods learned, but students are welcome to use their preferred software in solving lab assignments and evaluations. The techniques and principles learned in this class apply to any data analytics methods and are independent from the software package selected.

Definition of a Credit Hour

Success in this 3 credit hour course is based on the expectation that students will spend, for each unit of credit, a minimum of 45 hours over the length of the course (normally 3 hours per unit per week with 1 of the hours used for lecture) for instruction and preparation/studying or course related activities for a total of 135 hours.

Go to 4.08 / Definition and Assignment of Credit Hours for the policy and examples for different types of courses and credit hour offerings.

Measurable Student Learning Outcomes

Upon successful completion of this course, students will be able to:

i. Understand and appropriately use data analytics vocabulary.

ii. Recognize data analytics tools appropriate for specific datasets and analysis questions.

iii. Use data visualization techniques describe a dataset and to generate and answer specific analysis questions

iv. Formulate, implement, and interpret a regression analysis given a dataset and the software of choice.

v. Formulate, implement, and interpret a classification analysis given a dataset and the software of choice.

vi. Formulate, implement, and interpret a clustering analysis given a dataset and the software of choice.

- Self-learn new data analysis methods and their implementation in the preferred software package.

Required Texts/Readings Textbook

- Tan, P. N. (2019). Introduction to data mining, second edition. Pearson. New York, NY. Available in the WSU Bookstore and in Amazon.com. Select chapters are available on the author's website.

Other Readings

- Additional required reading materials will be specified on the corresponding lecture slides.

Other Equipment/Materials

Students are required to have complete access to a functioning laptop or PC with **internet capabilities**. This laptop or desktop computer must have the following **hardware**:

- Functioning **speakers**
- Functioning **microphone**
- Functioning **webcam**

The laptop or desktop computer must have the following **software**:

- (Required) The capability to connect to **Microsoft Teams** at designated times for office hours and contingency lectures and labs.
- (Required) **A text editor that allows conversion to pdf format** (since all submissions are needed in pdf format). Microsoft Word has this capability. If you do not have Microsoft Word on your PC or Laptop, Wichita State does provide free access to Microsoft Office 365 for students. Follow the instructions below to get Microsoft Office:
    1. Log in to MyWSU
    2. Click on Office 365 located on the "Home" tab
    3. Follow the Office 365 wizard instructions
- (Required) **Tableau** is a commercial data visualization software that supports the development of interactive visualizations and dashboards. Tableau has limited data analytics capabilities. Tableau Desktop is free for students and instructors around the world. Students can apply for a free license here.
- (Required) **RStudio** is an open-source software package that can be updated constantly to include packages for statistics, visualization, and other analysis tasks. It is free to anyone and can be downloaded here.

- (Required) **The preferred data analytics software** that supports ALL the analysis techniques covered in this course.

Before you begin your coursework, ensure that your computer meets technical standards (software, computer equipment, general skills, program management skills, communication skills, and managing your WSU e-mail) for use in this class.

Class Protocol and Conduct

This is a university course with commensurate expectations. We expect professional conduct in the classroom. Unprofessional conduct, e.g. off topic activities, may affect your grade or even your academic career. Be respectful at all times. Students demonstrating confrontational, disruptive, or threatening behavior will be asked to leave the classroom immediately and will not be allowed to return to the classroom for the remainder of the class period. Consequences of this behavior may also include (and are not limited to): Suspension from class for a minimum of one additional class period and report or referral to the WSU police department, Student Conduct and the WSU Care Team.

Below is a list of policies specific to this course:

1. Students enrolled in the course must be available to meet at the time of the class period for all the sessions of the term.
2. Each student is responsible for the syllabus, material discussed, distributed, or assigned during class, and all communications posted on blackboard. Students are responsible for attending lectures, preparing for flipped labs, participating in discussion forums, asking questions, and uploading lab assignments on time.
3. **Labs are flipped**. Students are expected to apply each method learned to a dataset of their choice. Lab instructions and preparation materials (if any) will be posted on Blackboard for students to practice before to the lab. The nature of the preparation materials will evolve throughout the semester. Initially, videos with detailed instructions will be posted. Then, only the code will be posted. Finally, students are expected to search for available online guides on how to implement the methods learned in the preferred software.
   3.1. The following will be graded during lab sessions:
      a. Students may choose to present their (preliminary) implementation results and conclusions to the class for feedback. Requires a draft of the project in a text editor. Only students presenting will get the extra credit points. Students wishing to present should contact the instructor 24 hours before the lab with the draft report for approval.
      b. Individual students' ability to answer questions about: the lab requirements, the recommended materials, and attempted implementation on their own datasets. *Each student should be prepared for labs and be able to answer questions without help from other team members.*
      c. Students not attending labs will receive a grade of 0 in the individual lab grade.
   3.2. A lab report will be submitted through Blackboard every week.
4. Students are encouraged to ask questions that help clarify concepts and problem-

solving during lectures, lab sessions, or office hours before quizzes and exams and before labs are due.

5. Outside articles presented during class sessions may be included in tests. Students must read them and ask questions before the tests.
6. Students must follow instructions on collaboration during quizzes and exams. Unless otherwise noted, quizzes are individual. Collaboration during quizzes is not allowed.
7. Students are not allowed to talk to each other during quizzes and exams. Regardless of the conversation topic, any instance of talking during an in-person exam will be considered a violation of academic integrity and students involved will receive a grade of 0 in the quiz or exam.
8. The instructor will notify the students of any resources they may access during evaluations.

It is the students' responsibility to check their grades regularly on Blackboard. Grades may only be revised within a week of grade announcements. If you think your exam quiz or lab report was graded unfairly, you must submit a written request to the instructor explicitly articulating your rationale for unfair/mistaken grading. You must do this within one week of the day the grade of the exam/quiz/report is posted on Blackboard. Note that the instructor will re-grade the entire exam/quiz.

Contact Policy

Email communication is always preferred. Feel free to email me any questions or concerns following these guidelines:

- Always use the course name in the subject line of the email
- Remember to sign your name.
- Always email me from your WSU email address. Email sent from personal email servers like Gmail, Yahoo, etc., have a tendency to end up in my spam folder, and I never see them. You may also email me through Blackboard via the Email My Instructor tab.
- Lab related e-mail questions must be sent more than 2 days (Monday through Friday) before the due date to ensure a timely response.
- Whenever you have a question about RStudio, please send as much detail as possible. If you want my help in understanding an error message in RStudio, include a print of your RStudio screen showing the code being used. Note that questions regarding the use of software are difficult to answer through e-mail. Keep in mind that there is a large community of R users online, and that if you search online for the error or the function you are trying to use, you will find many other users experiencing the same, along with possible solutions.
- You should NOT contact me for tech support.
    - Any technical problems involving your computer, or issues regarding file uploading or sharing, should go through OneStop. You can contact them at 316-978-3909. You can also fill out a request for help form at their website.
    - However, if you have a problem with accessing or uploading assignments, you *should* let me know before your assignment is due. You will also have to accompany this notification with the file in question, so I can verify that it is completed by the due date/time.

Response Time

- To Email: within 24 hours on Monday-Friday, slower during weekends and holidays.
- Feedback on Assignments and in-person test grades: within one week of submission. Students not attending the session when the exam was returned may ask for their paper during in-person office hours.

Academic Integrity

Students at Wichita State University are expected to uphold high academic standards. WSU will not tolerate a lack of academic integrity. Students are responsible for knowing and following the Student Academic Integrity Policy 2.17 (https://www.wichita.edu/about/policy/ch_02/ch2_17.php).

When the faculty member determines sanctions are warranted for violations of academic integrity, regardless of severity, the faculty member must report the infraction to the Office of Student Conduct and Community Standards. If you need more information about the process or wish to appeal a decision, please visit https://www.wichita.edu/about/student_conduct/ai.php

Please note in particular that *Unauthorized Use or Possession of Materials or Resources* (Policy item IV.B.2) includes unauthorized use of generative AI like ChatGPT or GPT4. In this course, all use of generative AI is prohibited except as may be specified in assignment instructions. Any student suspected of unauthorized use of generative AI may be asked to demonstrate their mastery of the assignment learning outcomes in an alternate format, such as a verbal interview or an additional problem set, before a determination is made as to whether an academic integrity violation report is warranted.

In this course, if the professor (or his designate) determines a student committed a violation of academic integrity in an exam, quiz, or assignment, then the grade on the assignment will be zero. Additionally, the professor may assign an 'F' as the course grade for repeated violations.

Sexual Harassment, Discrimination and Retaliation

Wichita State University is committed to creating a safe and healthy environment for all of our community members. To that end, we are partnering with Everfi to provide sexual assault prevention trainings that will help us all be informed on topics vital to achieving our goal. These trainings, in addition to being highly interactive and engaging, are based in research around the best practices for healthy communities. All students, staff and faculty are required to complete these trainings annually. Students will need to complete Everfi training in the fall semester before being able to enroll for the spring.

We are committed to the elimination of sexual misconduct, relationship violence, and stalking within the University community. These incidents may interfere with or limit an individual's ability to benefit from or fully participate in the University's educational programs. Students are asked to immediately report incidents to the University Police Department, (316) 978-3450 or students may contact the Title IX Coordinator at (316)

978-5177 or submit a report to the Office of Civil Rights, Title IX & ADA Compliance at Office of Civil Rights, Title IX & ADA Compliance (CTAC) Report Form (maxient.com).

Students may also report incidents to an instructor, faculty, or staff member, who are required by law to notify the Title IX Coordinator. If a student wishes to keep the information confidential, the student may speak with staff members of the Counseling and Psychological Services (316) 978-3440 or Student Health Services (316) 978-3620. For more information, please refer to policy 3.06 / Sexual Harassment, Discrimination and Retaliation for Employees, Students and Visitors.

Students with Disabilities

A disability is something that affects a major life activity. These life activities include but are not limited to, learning, walking, breathing, hearing, and seeing, in addition to many other physical, sensory functions, and psychological disabilities.

If you are a student with a disability or believe you might have a disability, which requires accommodations, please contact the Office of Disability Services (ODS) at www.wichita.edu/ods to discuss reasonable and appropriate accommodations and eligibility requirements. It is the University's goal that learning experiences be as accessible as possible. If you anticipate or experience physical or academic barriers based on disability ODS will review your concerns and determine, with you, what academic accommodations are necessary and appropriate for you. For example, adaptions of teaching methods, class materials or testing may be made on a case-by-case basis if warranted, as required by the Americans with Disabilities Act (ADA). All information and documentation of your disability is confidential and will not be released by ODS without your written permission.

The Office of Disability Services (DS) is located in Grace Wilkie, room 203, (316) 978-3309 (voice/tty) (316-854-3032 videophone). DS will review your concerns and determine, with you, what academic accommodations are necessary and appropriate for you. All information and documentation of your disability is confidential and will not be released by DS without your written permission.

Respect for Diversity

Wichita State University is committed to being an inclusive campus that reflects the evolving diversity of society. To further that goal, Wichita State University does not discriminate in its employment practices, educational programs, or activities on the basis of age (40 years or older), ancestry, color, disability, gender, gender expression, gender identity, genetic information, marital status, national origin, political affiliation, pregnancy, race, religion, sex, sexual orientation, or status as a veteran. Retaliation against an individual filing or cooperating in a complaint process is also prohibited.

Students from all diverse backgrounds and perspectives are welcome in this course and the diversity that students bring to this course should be viewed as a resource, strength, and benefit. All materials and activities are presented with the intent to be respectful of diversity: gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally or for other students or student groups. In addition, if any of our class meetings conflict with

your religious events, please let me know so that we can make arrangements for you.

Complaints or concerns related to alleged discrimination may be directed to the Office of Civil Rights, Title IX & ADA Compliance (CTAC), Wichita State University, telephone 316-978-3187.

Names & Pronouns

Everyone has the right to be addressed as and referred to by the name and pronouns (including non-binary pronouns) that they choose and that correspond to their gender identity. A student's chosen name and pronouns are to be respected at all times in the classroom.

For information about chosen name, students should visit www.wichita.edu/name. To provide preferred pronouns, click on "View and Update Personal Information" link in the Student Tools channel on the myWSU Home tab. Use the "Edit" buttons to update information.

Syllabus Policies and Student Resources

All students should familiarize themselves with the course-related policies and student resources that can be found at: www.wichita.edu/syllabuspolicies

These include, but may not be limited to:

- Important Academic Dates
- Proctoring exam options: For more information and requirements for using proctoring, please see https://wichita.edu/onlineproctoring
- Video and Audio recording
- Shocker Alert System
- Intellectual Property
- CARE Team
- Counseling and Prevention Services
- Student Health Services
- Heskett Center and Campus Recreation
- First Generation Students

Grading Scale

WSU uses a +/- grading scale for final grades and to calculate grade point averages. In this class, grades are assigned according to the following chart. (Other classes might assign grades differently: Be sure to understand the different grading scales in all your classes.)

Grading scale in the next page.

Grading Scale, continued

Points /Percentage	Letter Grade	Grade Points	Interpretation
≥ 92%	A	4.00	A range denotes excellent performance
87% - 92%	A-	3.70	
83% - 87%	B+	3.30	
80% - 83%	B	3.00	B range denotes good performance
77% - 80%	B-	2.70	
73% - 77%	C+	2.30	
70% - 73%	C	2.00	C range denotes satisfactory performance
67% - 70%	C-	1.70	
63% - 67%	D+	1.30	
60% - 63%	D	1.00	D range denotes unsatisfactory performance
57% - 60%	D-	0.70	
< 57%	F	0.00	

Assignments

Each topic is associated with a Lab assignment to apply the technique learned in a dataset using the preferred software. The lab component of the course is flipped. Self-learning of software is an important component of the course. Students should search for available resources online to learn how to implement the techniques learned in the preferred software package. Students will be provided with class time to ask and answer questions about the assignment prior to submission of their report. A lab report should be submitted through blackboard with the required contents by the specified due dates (see course schedule for dates).

Late Assignments

All class assignments will be turned in electronically through Blackboard. Blackboard assignments will not be available after the due date/time. Therefore, students must make sure to submit early enough to avoid technical issues. Technical difficulties will not be accepted as valid excuses for late submission, particularly within an hour or after the submission deadline. A grade of 0 will be given to missed assignments. If a student is having difficulty submitting their assignment through blackboard, they should send the assignment as an e-mail to the instructor before the due date/time, so that the assignment can be counted. Emails sent after the due date/time will be considered missed.

Missed Assignments and Exams

There are no make-up assignments, quizzes, or exams. Optional assignments will be used to replace the lowest assignment grades or missed assignments. In addition, the lowest quiz grade and the lowest exam grade will be dropped to accommodate students with special circumstances out of their control such as sickness, family emergency, military duty, jury duty, athletic events, conference presentations, job interviews, subpoena, religious observance, and others. Documentation of such events must be provided to the instructor before or within one week of the quiz for which the policy is to be implemented.

Undergraduate vs. Graduate Credit

Undergraduate students enrolled in 700 level courses will receive undergraduate credit (not graduate credit) unless they have a previously approved senior rule application or dual/accelerated enrollment form on file in the Graduate School. Undergraduate credit earned in 700 level courses cannot later be counted toward a graduate degree.

Extra Credit

Extra credit opportunities will be provided to all students throughout the semester in quizzes, exams, and labs. Students should take advantage of extra credit opportunities when they are available. Individual extra credit for students to increase failing grades at the end of the semester are not available.

Grading

Exams (average)	60%
Quizzes	20%
Group Lab Grade	10%
Individual Lab Grade	10%

Concealed Carry Policy

The Kansas Legislature has legalized concealed carry on public university campuses. Guns must be out of view, concealed either on the body of the carrier, or backpack, purse or bag that remains under the immediate control of the carrier.

Gun owners must familiarize themselves with WSU policy 11.19 / Weapons on University Property and the Kansas Board of Regent's weapons policy. If you believe that there has been a violation of these policies, please contact the University Police Department at (316) 978-3450.

Concealed Carry for lab courses or other courses where student belongings are unattended and out of immediate control: *Under the Concealed Carry Policy, a backpack or other bag used to carry a handgun must be within the immediate control of the individual. This course requires students to leave belongings such as backpacks or other bags out of reach and unattended for the duration of class time. Students who choose to carry a concealed handgun in a backpack or other bag must plan each day accordingly and are responsible for making alternate arrangements as necessary. Wichita State does not provide publicly available secure storage for concealed handguns.*

	Date	Class Topic	Readings	Flipped Lab Prep	Lab Report Due at Class Start Time Through Blackboard
Week 1	Tue, 8/22	Lecture: Introduction to class and data mining	Chapter 1	Installing R and Tableau on personal computer, explore datasets to use during the course.	
	Thu, 8/24	Lecture: Data and Preprocessing	Chapter 2; References on slides.	Select dataset and start exploring it.	
Week 2	Tue, 8/29	Lecture: Exploring data	References on slides. IME 734 Chart suggestion matrix summary	Implement data exploration techniques	
	Thu, 8/31	Lecture: Principles of data visualization (and exploring data)	Data visualization: basic principles	Ensuring principles of visualization in data exploration excercises	INDIVIDUAL Lab: Dataset Selection
Week 3	Tue, 9/05	Flipped Lab: (& Quiz) Principles of data visualization (and exploring data)			
	Thu, 9/07	Lecture: Introduction to Regression and Linear regression	Any regression analysis textbook, references on slides.	Define and implement linear regression in the preferred software.	INDIVIDUAL Lab: Exploring Data & Visualization
Week 4	Tue, 9/12	Flipped Lab: (& Quiz) Introduction to Regression and Linear regression			
	Thu, 9/14	Lecture: Introduction to Logistic regression	Book Section 4.6, References on Slides	Define and implement a logistic regression in the preferred software	Lab: Regression
Week 5	Tue, 9/19	**Exam: Midterm 1**			
	Thu, 9/21	Lecture: Introduction to Classification using Logistic Regression and Evaluating the Performance of a Classifier	Book sections: 3.2, 4.6, 3.4-3.6	Define and implement a logistic regression classification using the preferred software	

	Date	Class Topic	Readings	Flipped Lab Prep	Lab Report Due at Class Start Time Through Blackboard
Week 6	Tue, 9/26	Flipped Lab: (& Quiz) Introduction to Classification using Logistic Regression and Evaluating the Performance of a Classifier			
	Thu, 9/28	Lecture: Introduction to Cluster analysis and evaluation with k-means	Book Sections: 2.4, 7.1, 7.2, 7.5	Implement / evaluate a clustering analysis using K means	Lab: Intro to Classification with Logistic Regression
Week 7	Tue, 10/03	Flipped Lab: (& Quiz) Introduction to Cluster analysis and evaluation with k-means			
	Thu, 10/05	Lecture: Introduction to Association Analysis	Chapter 5	Implementing an association analysis	Lab: Intro to Clustering
Week 8	Tue, 10/10	Flipped Lab: (& Quiz) Introduction to Association Analysis			
	Thu, 10/12	Lecture: Data preprocessing, again!	Book section: 2.3, 4.11, References on slides.		Lab: Intro to Association Analysis
Week 9	Tue, 10/17	*NO CLASSES: FALL BREAK*			
	Thu, 10/19	Lecture: Classification and Regression Trees (CART)	Book section: 3.1, 3.2, 3.3, 3.4, 3.5, 3.6	Implement CART analysis	(Optional) INDIVIDUAL Lab: Data Preprocessing Revisited
Week 10	Tue, 10/24	Flipped Lab: (& Quiz) CART			
	Thu, 10/26	**Exam: MIDTERM 2**			Lab: Decision Trees
Week 11	Tue, 10/31	Lecture: Introduction to Ensemble methods	Book section: 4.10		
	Thu, 11/02	Lecture: Rule-based classifiers and Nearest Neighbor Classifiers	Book section: 4.2, 4.3	Implement rule based classifiers and KNN classifiers	

	Date	Class Topic	Readings	Flipped Lab Prep	Lab Report Due at Class Start Time Through Blackboard
Week 12	Tue, 11/07	Flipped Lab: (& Quiz) Rule-based classifiers and Nearest Neighbor Classifiers			
	Thu, 11/09	Lecture: Bayesian classifiers: Naïve Bayes and Bayesian Networks	Book Sections: 4.4-4.5	Implement bayesian classification analysis using the preferred software.	Lab: Rule-based and KNN classifiers
Week 13	Tue, 11/14	Flipped Lab: (& Quiz) Bayesian classifiers: Naïve Bayes and Bayesian Networks			
	Thu, 11/16	Lecture: Introduction to Artificial Neural Networks	Book section: 4.7	Implement ANN analysis using the preferred method	Lab: Bayesian Classifiers
Week 14	Tue, 11/21	Flipped Lab: (& Quiz) Introduction to Artificial Neural Networks			
	Thu, 11/23	*NO CLASSES: THANKSGIVING BREAK*	Book Section: 7.3	Implement hierarchical clustering using the preferred software.	INDIVIDUAL Lab (Optional): Introduction to Artificial Neural Networks
Week 15	Tue, 11/28	**Exam: MIDTERM 3**			
	Thu, 11/30	Lecture: Hierarchical clustering	Book Section: 7.3	Implement hierarchical clustering using the preferred software.	
Week 16	Tue, 12/05	Lecture: Quiz & Flipped Lab: Hierarchical clustering	Book section: 7.4	Implement density-based clustering.	
	Thu, 12/07	Lecture: Density-based clustering	Book section: 7.4	Implement density-based clustering.	Lab: Hierarchical clustering.
	TUESDAY Dec 12, 3:00 - 4:50 PM			**Exam: Final (see final exam matrix)** https://www.wichita.edu/services/registrar/final_examinations_schedule.php	