

MIT 5742
Data Science and Analytics
Fall 2022, Mod 2

Class time: Wednesday, 6:00pm-9:30pm (except November 11, 2022, 6 - 9:30 pm)

Location: HSC - 865 Research Parkway, Room 0357

Instructor:

Heshan Sun, Ph.D.
Professor, Richard Van Horn Professor of IT and Analytics
Price College of Business
Adams Hall, 3241
sunh@ou.edu
(405)325-5721 (phone)
Course website: <https://canvas.ou.edu/>

Office Hours:

Monday: 7:00pm-8:00pm (on Zoom)
Wednesday, 5:00pm-6:00pm (in person, faculty room near to Room 0357)
Or appointment by email (sunh@ou.edu)

Teaching Assistant :

Xinhui Zhan (email: xzhan@ou.edu)
Office hours: Tuesday/Thursday 7:00pm-8:00pm (on Zoom)
Zoom location: Available on Canvas

Course Description:

We are in the “Big Data” era. There are huge amounts of data available to business organizations today. Organizations that can harness this data, to discover meaningful patterns to make decisions, will be better positioned to compete in the marketplace. Analytics is a new field that facilitates this process, by making use of data, information technology, quantitative and statistical methods to identify patterns in data and develop predictions on potential events. Looking for patterns in data is not altogether a new activity. Scientists collect data and discover patterns using a host of scientific methods and tools. There are indeed a lot of common elements between data science approaches and analytics methods. However, while the former relies more on hypothesizing expected relationships, and collecting data that support/negate these expectations, the latter relies more on discovery of totally unexpected patterns, using techniques such as data mining. Hence, to gain new insights from large amounts of data that is available today, knowledge of traditional methods of data science as well as an understanding of newer analytic tools is necessary. In this course, we will review techniques that we have used quite often in data science, and then spend time in learning new methods in analytics. We will conduct data mining exercises and develop predictive models based on observed patterns in data. Through short lectures and plenty of hands-on exercises as pedagogical tools, course sessions will help us understand how organizations could identify patterns in data and leverage them for competitive strategies.

Objectives:

Upon completion of this course, students will:

- obtain understanding of data science and analytics
- learn specific data analytics topics such as decision trees, regression, cluster analysis, association rules, etc.
- get familiar with the data-mining procedure
- get familiar with various tools for data analytics

Course Materials

(Required) Business Analytics, by James R. Evans, Pearson Publishing. 3rd edition, ISBN-10: 0135231671. We only need chapters 1, 8, and 10 from this textbook.

(Optional) I benefit from the following books in preparation of this course. They are good books. I recommend you have them as well if you want to have more in-depth knowledge about data analytics and data mining.

1. Foster Provost & Tom Fawcett, Data Science for Business. O'Reilly Media. ISBN-10: 1449361323 (this book has interesting business examples for data science)
2. Data Mining: Practical Machine Learning Tools and Techniques, (3rd or 4th edition). by Ian Witten et al. (this book is a bit more technical and focus on machine learning algorithms)
3. Al Sweigart, Automate the Boring Stuff with Python, 2015, No Starch Press, ISBN-10: 1593275994. A free online edition of the textbook is available under a Creative Commons license:
<https://automatetheboringstuff.com>

Other reading materials will be made available in Canvas.

Technical Requirements

You need (1) a laptop, and (2) wireless Internet connection in class. We will use laptops regularly in the classroom. Given the nature of this course, you should bring a well-functioning laptop with the following software installed to each class.

Software applications:

1. **Anaconda** will be used for our Python programming. Anaconda has Jupyter notebook and Python. So you only need to download Anaconda at:
 - o Windows: <https://www.anaconda.com/download/#windows>
 - o MacOS: <https://www.anaconda.com/download/#macos>
2. **Power BI**. We will use Power BI for the analytics/visualization lab. Details regarding how to download and use Power BI will be available in Canvas.
3. **Excel**.

Weka: Weka is an open-source GUI tool that includes a collection of machine learning algorithms for data mining tasks. You can download it here, depending on your operation system (Windows or Mac OS): <https://www.cs.waikato.ac.nz/ml/weka/downloading.html>

The details about how to access these software applications will be available in class.

4. Other additional materials will be made available in Canvas.

Syllabus Changes

This is a dynamic syllabus, meaning it may undergo change. Although all changes in the syllabus will be announced in class, it is the student's responsibility to review the syllabus for changes each week. The latest syllabus will be posted on Canvas.

Grading

Grading will be based on a standard scale: (A=90+, B=80+, C=70+, D=60+, F=60-)

Final Exam	30%
Group project	30% (1 project)
Quizzes	30% (3 quizzes, 10% each)
Participation	10%
Total	100%

Final Exam (30%)

A final exam will be administered in the last class of the course.

Group project (30%)

We have one group project on data mining. Each group consists of around 5/6 people. The group will work together on a self-identified topic using data mining techniques. The deliverables will be a report and in-class presentation. Details will be available in class.

Quizzes (30%, 3 quizzes, 10% each)

Three in-class quizzes will be administered. Each quiz takes 20 minutes.

Participations (10%)

Participation is essential for this course. Therefore, a significant factor in determining grades will be the degree to which you participate in class.

The Participation grade is computed as follows:

Attendance – 60%, Active participation in class activities – 30%, Professionalism – 10%.

10 points will be deducted for each absence, failure to participate in labs and other class activities, and deficiencies in professionalism (such as improper laptop use, talking in class, arriving late or leaving early, etc.). Only if a student is attending a university-sponsored event or having well-documented illness will be excused. To get a full participation score, you also need to participate in class activities and discussion very actively.

Bonus Credits

One or two bonus credits may be available for participation in research projects and/or the attendance of professional development events. Opportunities will be suggested throughout the semester.

Weekly Schedule (Tentative)

Note: I expect everyone to have read all assigned materials **prior to** class.

Date	Topic	Materials	Notes
Oct 19	- Course introduction - Background: data science and analytics	- Syllabus - Evans book: Chapter 1	
Oct 26	Introduction to business analytics - Lab on the data-mining approach	- Evans book: Chapter 1 (cont'd) - Data-mining methods (available on Canvas) - Data Types in Statistics (available on Canvas)	Group formed and group project available
Nov 2	- Decision trees and consumer behavior - Lab on decision trees	- Entropy: How Decision Trees Make Decisions	
Nov 9	- Regression - Lab on regression	Evans book: Chapter 8	Quiz 1: Decision trees
Nov 11 (Friday)	- Regression (cont'd) - Lab on regression	Evans book: Chapter 8	
Nov 16	- Logistic regression - Association rules	- Evans book: Chapter 10 (p.g. 370-372) - "Complete guide to Association Rules" - (Optional reading) Retail Case Study Example	Quiz 2: Linear regression

Nov 23	No class: Thanksgiving Holiday!		
Nov 30	<ul style="list-style-type: none"> - Cluster analysis - Lab on cluster analysis - Tableau, Power BI lab/demo 	(required) Evans book: Chapter 10 (p.g. 358-362). - (optional) "The Most Comprehensive Guide to K-Means Clustering You'll Ever Need" (Optional reading) - The predictive power of social media. - (Optional reading) Behavior-based customer segmentation for more effective retail marketing.	Quiz 3: Logistic regression and association rules
Dec 7	<ul style="list-style-type: none"> - Group project presentation - Final Exam study guide 		Group project due
Dec 14	- Final Exam		

Policies

University Policies and Resources

I expect everyone to abide the university's honor code:

"On my honor, as a student at the University of Oklahoma, I affirm that I will neither give nor receive inappropriate aid in the completion of any academic exercise. I understand that it is my responsibility to comply with the Academic Misconduct Code."

Reasonable Accommodation Policy:

"The University of Oklahoma is committed to providing reasonable accommodation for all students with disabilities. Students with disabilities who require accommodations in this course are requested to speak with the professor as early in the semester as possible. Students with disabilities must be registered with the Office of Disability Services prior to receiving accommodations in this course. The Office of Disability Services is located in Goddard Health Center, Suite 166, phone 405/325-3852 or TDD only 405/325-4173."

Religious Holiday Policy:

"It is the policy of the University to excuse the absences of students that result from religious observances and to provide without penalty for the rescheduling of examinations and additional required classwork that may fall on religious holidays."

OU Resources on Copyright and Piracy:

<http://www.ou.edu/content/ouit/security/copyright.html>

Adjustments for Pregnancy/Childbirth Related Issues:

Should you need modifications or adjustments to your course requirements because of documented pregnancy-related or childbirth-related issues, please contact me as soon as possible to discuss.

Generally, modifications will be made where medically necessary and similar in scope to accommodations based on temporary disability. Please see www.ou.edu/content/eoo/faqs/pregnancy-faqs.html for commonly asked questions.

Title IX Resources

For any concerns regarding gender-based discrimination, sexual harassment, sexual misconduct, stalking, or intimate partner violence, the University offers a variety of resources, including advocates on-

call 24/7, counseling services, mutual no contact orders, scheduling adjustments and disciplinary sanctions against the perpetrator. Please contact the Sexual Misconduct Office 405-325-2215 (8-5, M-F) or OU Advocates 405-615-0013 (24/7) to learn more or to report an incident. Additional information can be found at: <http://www.ou.edu/content/eoo.html>

Course Attendance Policy (Covid 19)

A temporary university policy has been established to protect the OU community by ensuring that students who are ill or required to isolate feel encouraged to remain at home. Missing a class session or other class activity due to illness or isolation will not result in a penalty for the absence, and the student will not be asked to provide formal documentation from a healthcare provider to excuse the absence. This policy is based on all students and faculty adhering to the principles of integrity, honesty, and concern for others.

Students who are experiencing symptoms of COVID-19, including cough, fever, shortness of breath, muscle pain, headache, chills, sore throat, loss of taste or smell, congestion or runny nose, nausea or vomiting, or diarrhea or who have been in close contact with others who have symptoms should:

- Remain at home to protect others
- Ensure that any needed screening has been conducted (COVID-19 Screening and Reporting Tool) and any needed treatment obtained
- Contact the instructor prior to absence or inability to participate, if possible, and provide an honest report of the reason for which you cannot attend class or complete a course activity
- Continue to complete coursework to the extent possible, using Canvas, zoom, and other online tools
- Submit assignments electronically to the extent possible and as directed by the instructor
- Communicate with the instructor to arrange modifications to deadlines or work requirements or reschedule exams or other important course activities, when it is necessary.