Fundamentals of Data Science for Engineers (SIE 433/533)
Tue/Thu 12:30 – 1:45PM, Old Engineering 301 & **In-Person**

Instructor and Contact Information

Instructor:	Dr. Jian Liu, Associate Professor of the Department of Systems & Industrial Engineering
Office:	ENGR 221
Email:	jianliu@email.arizona.edu
Phone:	520-621-6548
Office Hours:	Tue. 3:30 – 4:30 PM, or by Appointment (OBA)
Office Hour for Online Students:	Tue. 8 – 9 PM, by D2L –> UA Tools –>Zoom
Teaching Assistants (TAs):	Yinwei Zhang (zhangyinwei@email.arizona.edu) Xi Chen (xic@email.arizona.edu)
Office Hours:	Mon. /Wed. 10:30 – 11:30 AM, Xi Chen, OBA
Office Hours for Online Students:	Mon/Wed. 8:00 – 9:00 PM, Yinwei Zhang, OBA

All office hours will be held via Zoom: D2L –> UA Tools –>Zoom

This class is scheduled to be taught in the **IN-PERSON** modality and available for distance learning sections. All the students have access to Panopto lecture recordings.

Course Description and Overview

This course will provide senior undergraduate and graduate students from a diverse engineering disciplines with fundamental concepts, principles and tools to extract and generalize knowledge from data. Students will acquire an integrated set of skills spanning data processing, statistics and machine learning, along with a good understanding of the synthesis of these skills and their applications to solving problem. The course is composed of a systematic introduction of the fundamental topics of data science study, including: (1) principles of data processing and representation, (2) theoretical basis and advances in data science, (3) modeling and algorithms, and (4) evaluation mechanisms. The emphasis in the treatment of these topics will be given to the breadth, rather than the depth. Real-world engineering problems and data will be used as examples to illustrate and demonstrate the advantages and disadvantages of different algorithms and compare their effectiveness as well as efficiency, and help students to understand and identify the circumstances under which the algorithms are most appropriate.

Course Objectives and Expected Learning Outcomes

The objective of this course is to prepare engineering students with fundamental data science training. While mathematical methods and theoretical aspects will be covered, the primary objective of this course is to equip the students with the tools and principles needed to solve both the traditional and the novel problems in data science practice, including data integration, exploratory analysis, modeling, evaluation and effective communication. Students who take this course will get familiarized with a broad cross-section of models and algorithms, with the emphasis given to their applications in problem solving.
By taking this course, both **undergraduate** and **graduate** students will possess the capability to:

(i) conduct data integration and preprocessing;
(ii) understand a wide variety of learning algorithms; and

(iii) appropriately apply the algorithms to real-world engineering problems.

For **graduate** Students, additional learning outcomes include:

(i) theoretical underpinnings of the learning algorithms, and

(ii) approaches of evaluating the learning outcomes and their theoretical justifications.

The instructor and the TAs will be available to answer questions and/or discuss course materials during the office hours or by appointment. Office hours for online students will be offered separately through online meeting system at specified time. Extra office hours may be offered according to instructors' availability. Email communication is preferred when the instructor is on travel. Response time to email questions from instructors will be within 24 hours.

Course Prerequisites

For *undergraduate* students, the prerequisite course is: SIE 305 (Introduction to Engineering Probability and Statistics), equivalent courses or consent of the instructor. For Advanced Standing of undergraduate students, please visit the webpage for detail information to obtain the advanced standing: http://sie.engr.arizona.edu/advanced-standing. For *graduate* students, the prerequisite course is SIE530 (Engineering Statistics), equivalent courses (such as SIE 500A taken in parallel with this course) or consent of the instructor. The prerequisite knowledge that is covered in SIE530 but not in SIE 305 will be reviewed for related topics, if applicable, for undergraduate and graduate students to have adequate background for learning new data science knowledge.

It is recommended that you have some background in R or Python programming languages (but this not required, we will have dedicated lectures, sample codes and office hours to help student with programming).

Temporary Lecture Schedule:

Lecture	Week	Topics	Course Work
1	1	Course overview & Introduction	
2		Data Structure & Representation	
3	2	Data Integration & Preprocessing with Python	
4		Data Similarity vs. Dissimilarity	HW 1
5	3	Data Transformation & Discretization	
6		Concepts and Principles for Computer to Learn	
7	4	Learning Procedure & Approximation-Generalization Tradeoff	HW 2
8		Supervised Learning - Linear Regression I	
9	5	Linear Regression II	
10		Supervised Learning - Logistic Regression I	HW 3 & Proposal
11	6	Logistic Regression II	
12		Link Functions & Linear Discriminant Analysis	HW 4
13	7	Overfitting & Cross Validation	
14		Model Selection & Information Criteria	
15	8	Review Session for Exam I &Project Discussion	
16		**Exam I (in class)**	**Exam I**
17	9	Margin Concept and Support Vector Machines	
18		Support Vector Machines	HW 5
19	10	Kernel Methods	
20		Naïve Bayes Method	

21	11	Cross Validation & Bootstrap	
22		K-Nearest Neighbors Method & K-Means Method	HW 6
23	12	Principal Component Analysis	
24		Neural Networks	
25	13	Decision Tree	HW 7
26		Project Preparation/Presentations	
27	14	Project Preparation/Presentations	
28		Review Session for Exam II	Presentation & report
29	15	**Exam II**	**Exam II**

The above topics and schedule are subject to change. Revisions in the syllabus may occur as the semester progresses.

Reading Materials

Main: Lecture notes, provided and can be downloaded from D2L course website

Recommended reference books:

Pattern Recognition and Machine Learning - by C. M. Bishop, Springer 2006.

Machine Learning - by Tom M. Mitchell, McGraw-Hill, 1997

The Elements of Statistical Learning - by T. Hastie, R. Tibshirani, and J. Friedman, 2009

Papers and reports that instructor uploaded online for this course.

Grading Scale and Grade Policy

I - Grading:

Course grades for **undergraduate** section will be determined based on the following items:

Course Work	Points	Percentage
Homework	25	25%
Exam I	25	25%
Exam II	25	25%
Final Project Team Proposal (1 page) Team Presentation (15 min) Team Report (5 pages)	25	25%
TOTAL		**100%**

Course grades for **graduate** section will be determined based on the following items:

Course Work	Points	Percentage
Homework	25	25%
Exam I	25	25%
Exam II	25	25%
Final Project Team Proposal (1 page) Team Presentation (15 min) Team Report (5 pages + evaluation/ justification section)	25	25%
TOTAL		**100%**

The grading scheme will follow the distribution below.

Points	Percentage	Letter Grade
90-100	90%-100%	**A**
80-89	80%-89%	**B**
70-79	70%-79%	**C**
60-69	60%-69%	**D**
<60	<60%	**E**

There is no extra credit for any student.

II - Test and Exam

The test/exam date and time are scheduled based on course's schedule and all students are required to participate in the tests/exams on time. The only exception is for medical reasons (if it is a medical reason, only Doctor's notes are acceptable; Others (e.g., purchasing pills by yourself) etc. are not acceptable).

Students should read the test/exam instruction carefully before it starts, whether it is open-book or a close-book test/exam. If it is an open-book test/exam, students who forget to bring a textbook or other materials, are not allowed to borrow book or other materials from other students in the exam room.

Graduate students who register SIE533 and undergraduate students who register SIE433 will work on different exam papers to reflect different requirements on graduate students and undergraduate students, and will be counted to the final grades.

Students with disability, please contact the TA to arrange your tests/exams or schedule with the Disability Resource Center (DRC). All students will take the exams through *Examity* or with proctors approved by the instructor.

Final exam policy will follow UA's policy http://registrar.arizona.edu/courses/final-examination-regulations-and-information?audience=students&cat1=10&cat2=31.

III - Homework (HW)

Homework will be assigned throughout the semester. All homework will be submitted on D2L before 11:59 PM on the due date (See the weekly schedule). Penalty will be given to late homework. The exceptions for illness and events may be considered (if it is a medical reason, please provide doctor's record or a signed letter by the doctor; Notes from advisors or Deans are needed for the participation of conferences.):

1) Homework submitted on the second day: -15% (15% of score of your homework will be taken off).
2) Homework submitted on the third day: -30%
3) Homework submitted on the fourth day or later: -100%

If students submit the homework elsewhere (e.g., put under the door of offices of instructor or TA etc.), the students have to take the responsibility for the missing of these HWs (i.e., receiving those late penalty above until the missing HW is found or resubmitted).

Note: all homework, if you have a reference section (e.g., HW1), must strictly follow the reference format we taught in the class.

IV – Course Project

Detailed course project introduction and requirements will be announced and specified in a separate document. Temporary course project guideline document is listed below:

SIE 433/533 Fundamentals of Data Science for Engineers

-- Course Project Guideline

1 - Objective: the objective of the course project is for students to specify and implement an engineering data analytics tools introduced in this course, and prove their functional effectiveness in solving a real-world problem. Through a series of activities on problem definition, solution formulation and algorithm implementation, oral presentation and report composing, student teams are expected to demonstrate their knowledge and understanding of the key subjects of engineering data analytics.

2 - Phases of the course project:

- *Problem definition*: in this phase, students are expected to explore and describe the problems of interest and provide some investigation of background and data availability. A review of literature in the selected area is preferable. A justifiable hypothesis should be formulated to reflect the objective of the proposed project work, which will be specified in an initial proposal. Students are recommended to combine this course project with the data analytics needs in their senior design, master report and dissertation research. Students may also find potential project ideas and data sets from the online database listed in the section of *Online Database*.
- *Solution formulation*: in this phase, students should try to map out a solution that is implementable with the available time and resources. It is desirable for students to justify the potential effectiveness of the proposed solution, according to the specific nature of the problem. It is acceptable to perform solution formulation and implementation (next phase) iteratively. However, there is merit in trying to think things through before implementing it.
- *Implementation*: in this phase, students are expected to implement the solution, or some aspects of the solution to demonstrate its feasibility and effectiveness. Hypothesis formulated in the problem definition phase should be tested and the findings should be assessed with reasonable metrics.

3 - Online Database:

- *Kaggle* (https://www.kaggle.com/): "is an online community of data scientists and machine learners, owned by Google, Inc. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges. "[1]
- *UCI Machine Learning Repository* (https://archive.ics.uci.edu/ml/datasets.html): "The UCI Machine Learning Repository is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms."[2]
- *Additional datasets* (https://github.com/awesomedata/awesome-public-datasets): the problems and datasets are self-explained.

4 - Structure of the Project: Students need to work in groups of 3 or 4 with the prior consent of the instructor. Distance-learning students may join a group or complete the project individually. Groups with both undergraduate students and graduate students are encouraged. Every group will be assigned

[1] https://en.wikipedia.org/wiki/Kaggle
[2] https://archive.ics.uci.edu/ml/about.html

with an ID for documents submission and grading. The project account for 25 points of final grade and is to be structured into the following three *project components*:

(i) *Initial Proposal* (**4 points**): a written proposal (limit: one letter-size page with optional appendix of data description) should be submitted to the D2L dropbox folder by 11:59PM, September 26, 2020. The proposal should present the problems to be explored, the background and motivations, objective(s), dataset characteristics, and the milestone progress to be expected.

(ii) *Presentation* (**11 points**): each project team will make a 15-minute group oral presentation, including 2-3 min Q&A. Each team member is required to present. All groups present to the instructor and TAs via Zoom. The presentations are arranged according to the course schedules. Additional time slots may be added.

(iii) Final Report (**10 points**): a final, concise, description of what was attempted, what was accomplished, and what was learned from the project. Ideally, this will be a revision of the proposal, changing the statement of proposed work to the descriptions of accomplished work, refining the goals and objectives, and presenting some final conclusions and findings. The page limit for the final report is **five** (5) letter-size pages. For **graduate** students, an appendix section titled "*Theoretical Justification and Evaluation*" ***must*** be included in the final report. A final report from a team with graduate student members without this appendix section ***will be returned without review***, and the points for the final report of the team will be 0. Optional appendix may be used to include data and results visualization, algorithms, flowcharts and python codes. The final report is due to D2L dropbox folder at 11:59PM, Sunday, December 13, 2020.

V – Teamwork Evaluation: project team members will be provided a *Teamwork Evaluation Form* to self-evaluate and evaluate team members' contributions to the corresponding project components, including proposal, presentation and report, with a score scale from 0 to 10 (0 = no contribution; 10 = greatest contribution). The individual's teamwork score for each project component will be calculated as the average evaluation scores from all team members, scaled by 10. For example, the evaluation scores of project proposal for student A given by all the five team members, including his/her self-evaluation score, are 10, 9, 8, 7, 6. Then, A's teamwork score is (10+9+8+7+6)/5/10 = 0.8. If the team receive a project proposal score of 3, A's adjusted proposal score will be 3x0.8 = 2.4. The final score (maximum points: 25) of a student will be calculated as the summation of his/her three adjusted project component scores.

Attentions for COVID-19 Pandemic

The use of masks will be required for everyone, regardless of vaccination status, in all indoor spaces where continuous physical distancing is not possible. Students who forget to bring a mask to class should be offered one. Masks and cleaning supplies will be available in every classroom. If a student refuses to wear a mask in a classroom or other designated space when reminded, they should be asked to leave the space. An online Incident Report Referral Form to the Dean of Students Office will be submitted.

The Disability Resource Center is available to explore face coverings and accessibility considerations if you believe that your disability or medical condition precludes you from utilizing any face covering or mask option. DRC will explore the range of potential options as well as remote course offerings. Should DRC determine an accommodation to this directive is reasonable, DRC will communicate this accommodation with your instructor.

Classroom attendance:

- If you feel sick, or may have been in contact with someone who is infectious, stay home. Except for seeking medical care, avoid contact with others and do not travel.
- Notify your instructor(s) if you will be missing a course meeting or an assignment deadline.
- Non-attendance for any reason does **not** guarantee an automatic extension of due date or rescheduling of examinations/assessments.
- Please communicate and coordinate any request directly with your instructor.
- If you must miss the equivalent of more than one week of class, you should contact the Dean of Students Office DOS-deanofstudents@email.arizona.edu to share documentation about the challenges you are facing.
- Voluntary, free, and convenient COVID-19 testing is available for students on Main Campus.
- COVID-19 vaccine is available for all students at Campus Health.
- Visit the UArizona COVID-19 page for regular updates.

Academic advising: If you have questions about your academic progress this semester, please reach out to your academic advisor (https://advising.arizona.edu/advisors/major). Contact the Advising Resource Center (https://advising.arizona.edu/) for all general advising questions and referral assistance. Call 520-626-8667 or email to advising@.arizona.edu.

Life challenges: If you are experiencing unexpected barriers to your success in your courses, please note the Dean of Students Office is a central support resource for all students and may be helpful. The Dean of Students Office can be reached at (520) 621-2057 or DOS-deanofstudents@email.arizona.edu.

Physical and mental-health challenges: If you are facing physical or mental health challenges this semester, please note that Campus Health provides quality medical and mental health care. For medical appointments, call (520) 621-9202. For After Hours care, call (520) 570-7898. For the Counseling & Psych Services (CAPS) 24/7 hotline, call (520) 621-3334.

Class Recordings: For lecture recordings, which are used at the discretion of the instructor, students must access content in D2L only. Students may not modify content or re-use content for any purpose other than personal educational reasons. All recordings are subject to government and university regulations. Therefore, students accessing unauthorized recordings or using them in a manner inconsistent with UArizona values and educational policies (Code of Academic Integrity and the Student Code of Conduct) are also subject to civil action.

Classroom Behavior Policy:

Mainly from the university classroom policy (adopted by the Faculty Senate):

- Not leaving early. Early leaving will distract both the instructor and students

- Not talking with other classmates while the instructor or another student is speaking. If a student has a question or comment, he or she should raise a hand, rather than starting a conversation about it with a neighbor
- Not packing backpacks to leave until the instructor has dismissed class
- Showing respect and concern for others by not monopolizing class discussion. Students must allow others time to give their input and ask questions. Students should not stray from the topic of class discussion
- Not eating and drinking during class time

Academic Integrity (*Very Important!)* and Code of Academic Integrity:

Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work/exercises must be the product of independent effort unless otherwise instructed. Students are expected to adhere to the UA Code of Academic Integrity as described in the UA General Catalog. See: http://deanofstudents.arizona.edu/academic-integrity/students/academic-integrity.

Both instructor and TA will carefully exam all of your homework, reports and exams to prevent plagiarism following: deanofstudents.arizona.edu/policies-and-codes/code-academic-integrity. For example:

No copy of other people's homework In grading your homework, TA will exam all of your homework carefully and catch anyone who is copying other people's homework. Even if they are from the same software's output, TA can still judge whether it is a copy of others or not based on your writing and formatting.

No copy or discussion in the tests/exams TA and other students all will report these behaviors in exams and your exam papers will also be checked by TA carefully for any cheating behavior. Do not seat too close to each other in the exam.

Teaching Assistant's Responsibility

TA is mainly responsible for assisting the instructor in various issues, including grading homework, grading exams, teaching part of software and tutorial, helping students in programming, helping students set up their websites and webpages, etc.

Send Feedback to US:

If you have any questions, suggestions or comments related to the class, you are very welcome to contact the instructor or TAs directly. We have several ways for communications:

1) In-Class Feedback Papers (White papers) (Anonymous)
2) Office Hours
3) Emails
4) Individual Appointment (if you cannot come in Office Hour).

Threatening Behavior Policy:

The UA Threatening Behavior by Students Policy prohibits threats of physical harm to any member of the University community, including to oneself. See: http://policy.arizona.edu/education-and-student-affairs/threatening-behavior-students.

UA Nondiscrimination and Anti-harassment Policy:

The University is committed to creating and maintaining an environment free of discrimination, see: http://policy.arizona.edu/human-resources/nondiscrimination-and-anti-harassment-policy.

Inclusive Excellence is a fundamental part of the University of Arizona's strategic plan and culture. As part of this initiative, the institution embraces and practices diversity and inclusiveness. These values are expected, respected and welcomed in this course.

This course supports elective gender pronoun use and self-identification; rosters indicating such choices will be updated throughout the semester, upon student request. As the course includes group work and in-class discussion, it is vitally important for us to create an educational environment of inclusion and mutual respect.

Additional Resources for Students Statement: Office of Diversity (http://diversity.arizona.edu/)
http://www.health.arizona.edu/counseling-and-psych-services
http://oasis.health.arizona.edu/hpps_oasis_program.htm

Accessibility and Accommodations (foe students with Disability)

At the University of Arizona we strive to make learning experiences as accessible as possible. If you anticipate or experience physical or academic barriers based on disability or pregnancy, you are welcome to let me know so that we can discuss options. You are also encouraged to contact Disability Resources (520-621-3268) to explore reasonable accommodation.

If you have reasonable accommodations, please plan to meet with me by appointment or during office hours to discuss accommodations and how my course requirements and activities may impact your ability to fully participate.

If our class meets at a campus location: Please be aware that the accessible table and chairs in this room should remain available for students who find that standard classroom seating is not usable.

Their website is http://drc.arizona.edu/instructors/syllabus-statement.

Subject to Change Statement

The information contained in the course syllabus, may be subject to change, as deemed appropriate by the instructor, see http://policy.arizona.edu/faculty-affairs-and-academics/course-syllabus-policy-undergraduate-template.