

INSY 5378: Data Science: A Programming Approach Spring 2017

Instructor: Gene Moo Lee, Ph.D.

Email: gene.lee@uta.edu

Profile: <http://www.uta.edu/profiles/gene-lee>

Time and Place of Class Meetings:

- Section 001: TR 5:30~6:50 PM, COBA 256
- Section 002: TR 2:00~3:20 PM, COBA 339

Office Number: COBA (Business Building) 516

Office Telephone Number: 817-272-3084

Office Hours: TR 3:30~5 PM or by appointment

Description of Course Content: The world is awash in data and companies are now trying to discern patterns and predict behaviors of both consumers and competitors to gain and sustain a competitive advantage. The unstructured nature of data as well as the myriad sources they come from make it particularly challenging for companies to systematically capture, cleanse, store, and analyze the data. Python is a simple yet powerful language that has a rich ecosystem to facilitate the analysis of such complex data. The aim of this course is to acquaint students with aspects of the Python language that are necessary to effectively function as a data scientist. Upon successful completion of the course, students will be familiar with data structures and programming constructs in the Python language, accessing data from files and databases, Social Network Analysis, Text Mining, and Deep Learning.

Student Learning Outcomes: The aim of this course is to acquaint students with aspects of the Python language that are necessary to effectively function as a “data scientist.” Upon successful completion of the course, students will be familiar with:

- a) Data structures and programming constructs in the Python language. Specifically, students will have a good grasp of lists, tuples, dictionaries, classes, selection (e.g., `if . . else`), and iteration (e.g., `while` and `for` loops).
- b) Accessing data from files (e.g., text, CSV, JSON, etc.).
- c) Text mining, topic modeling, and sentiment analysis using NLTK and Gensim
- d) Basics of Social Network Analysis using Networkx
- e) Machine learning algorithms using Scikit-learn
- f) Basics of Deep Learning using TensorFlow

Required Textbooks and Other Course Materials:

- **Textbook:** *Data Science from Scratch: First Principles with Python* by Joel Grus, O'Reilly Media, 2015, ISBN: 978-1491901427
- **(Optional) Supplementary Book:** *Introduction to Computing Using Python: An Application Development Focus, Second Edition* by Ljubomir Perkovic, John Wiley & Sons, Inc., 2015, ISBN: 978-1-118-89094-3 (e-Text ISBN: 978-1-118-89105-6)
- **Software:** Python 2.7 with IDE (preferably Anaconda from Continuum Analytics; it can be downloaded at <https://www.continuum.io/downloads>); other software packages as needed
- **Codes:** Dropbox Folder provided by Instructor <https://goo.gl/YeoXIH>

Descriptions of major assignments and examinations:

The distribution of points will be as follows:

Exam 1	20 points
Exam 2	20 points
Homeworks (2 sets)	10 points (5 points each)
Group Projects (2 sets)	20 points (10 points each)
Final Project Presentation	10 points
Final Project Report	20 points
Class participation/Quizzes	Extra credits (1 point for each)

Grading: The following criteria will be used to assess your grade (**no rounding!**):

A (≥ 90 points), B (≥ 80 points), C (≥ 70 points), D (≥ 60 points), F (< 60 points)

Attendance: At The University of Texas at Arlington, taking attendance is not required but attendance is a critical indicator in student success. Each faculty member is free to develop his or her own methods of evaluating students' academic performance, which includes establishing course-specific policies on attendance. As the instructor of this section, I will consider attendance mandatory for all lectures. If you miss a class, you are responsible for the materials covered. **Pop quizzes or class attendance** may be used towards Extra Credit upon the instructor's discretion.

However, while UT Arlington does not require instructors to take attendance in their courses, the U.S. Department of Education requires that the University have a mechanism in place to mark when Federal Student Aid recipients "begin attendance in a course." UT Arlington instructors will report when students begin attendance in a course as part of the final grading process. Specifically, when assigning a student a grade of F, faculty report the last date a student attended their class based on evidence such as a test, participation in a class project or presentation, or an engagement online via Blackboard. This date is reported to the Department of Education for federal financial aid recipients.

Exams: There will be 2 exams. You are responsible for everything that is covered in the classroom, including additional materials that the instructor may discuss in class. There are **no make-up exams**. Under extenuating circumstances (e.g., medical emergency, family emergency, work-related travel, etc.), the average score of Exam, Homeworks, and Projects will replace the missed exam score. You can only use this excuse for one exam. Note that there will be **no final exam**; instead there will be **final project**.

Homeworks and Group Projects: You will have 2 sets of individual homeworks and 2 sets of group projects. Homeworks should be conducted individually. All homeworks and projects must be submitted via BlackBoard by the due date and time specified by the instructor. Late submissions will receive a score of 0 (no exceptions).

* **Plagiarism:** Plagiarism issues are taken very seriously. It is allowed to discuss high-level ideas with classmates. However, copying code or data (either fully or partially) is considered as academic dishonesty. If you are not sure about the boundary, please contact the instructor.

Final Project: An important component of the course is a group project. The class will be divided into groups of three for this purpose. Each group will pick a topic, analyze a dataset, write up a report, and make a presentation to the class. The project is meant to reinforce concepts taught in the lectures, slides, and/or additional readings.

1. **Kick-off report (due 3/3):** Before starting the project, groups must submit a kick-off report containing (i) team members, (ii) problem statement, (iii) required datasets, and (iv) potential implications.
* Instructor's approval is needed before proceeding to work on the project.
2. **Presentation (4/27, 5/2, 5/4):** Each group will make a 10-minute presentation at the end of the semester. It is not necessary that all members speak in the presentation. Peer evaluation will be combined with instructor's score. Attendance will be taken.
3. **Final Report (due 5/9):** The report should include (i) problem statement, (ii) data and its summary description, (iii) model with justification, (iv) progress, (v) business implications; and (vi) lesson learned.
4. **Intra-group Peer Evaluation (due 5/9):** It is possible for members in the group to receive different scores based on individual contribution.

*** Note that the same groups will work on the group projects as well.**

Further details will be provided in class and/or Blackboard.

Expectations for Out-of-Class Study: Beyond the time required to attend each class meeting, students enrolled in this course should expect to spend at least additional 9 hours per week of their own time in course-related activities, including reading required materials, completing assignments, and preparing for exams/quizzes.

UNIVERSITY and COLLEGE POLICIES

Grade Grievances: Any appeal of a grade in this course must follow the procedures and deadlines for grade-related grievances as published in the current University Catalog.

For undergraduate courses, see <http://catalog.uta.edu/academicregulations/grades/#undergraduatetext>;

For student complaints, see <http://www.uta.edu/deanofstudents/student-complaints/index.php>.

Drop Policy: Students may drop or swap (adding and dropping a class concurrently) classes through self-service in MyMav from the beginning of the registration period through the late registration period. After the late registration period, students must see their academic advisor to drop a class or withdraw. Undeclared students must see an advisor in the University Advising Center. Drops can continue through a point two-thirds of the way through the term or session. It is the student's responsibility to officially withdraw if they do not plan to attend after registering. **Students will not be automatically dropped for non-attendance.** Repayment of certain types of financial aid administered through the University may be required as the result of dropping classes or withdrawing. For more information, contact the Office of Financial Aid and Scholarships (<http://wweb.uta.edu/aao/fao/>).

Disability Accommodations: UT Arlington is on record as being committed to both the spirit and letter of all federal equal opportunity legislation, including *The Americans with Disabilities Act (ADA)*, *The Americans with Disabilities Amendments Act (ADAAA)*, and *Section 504 of the Rehabilitation Act*. All instructors at UT Arlington are required by law to provide "reasonable accommodations" to students with disabilities, so as not to discriminate on the basis of disability. Students are responsible for providing the instructor with official notification in the form of **a letter certified** by the Office for Students with Disabilities (OSD). Only those students who have officially documented a need for an accommodation will have their request honored. Students experiencing a range of conditions (Physical, Learning, Chronic Health, Mental Health, and Sensory) that may cause diminished academic performance or other barriers to learning may seek services and/or accommodations by contacting:

The Office for Students with Disabilities, (OSD) www.uta.edu/disability or calling 817-272-3364.

Information regarding diagnostic criteria and policies for obtaining disability-based academic accommodations can be found at www.uta.edu/disability.

Counseling and Psychological Services, (CAPS) www.uta.edu/caps/ or calling 817-272-3671 is also available to all students to help increase their understanding of personal issues, address mental and behavioral health problems and make positive changes in their lives.

Non-Discrimination Policy: *The University of Texas at Arlington does not discriminate on the basis of race, color, national origin, religion, age, gender, sexual orientation, disabilities, genetic information, and/or veteran status in its educational programs or activities it operates. For more information, visit uta.edu/eos.*

Title IX Policy: The University of Texas at Arlington ("University") is committed to maintaining a learning and working environment that is free from discrimination based on sex in accordance with Title IX of the Higher Education Amendments of 1972 (Title IX), which prohibits discrimination on the basis of sex in educational programs or activities; Title VII of the Civil Rights Act of 1964 (Title VII), which prohibits sex discrimination in employment; and the Campus Sexual Violence Elimination Act (SaVE Act). Sexual misconduct is a form of sex discrimination and will not be tolerated. *For information regarding Title IX, visit www.uta.edu/titleIX or contact Ms. Jean Hood, Vice President and Title IX Coordinator at (817) 272-7091 or jmhood@uta.edu.*

Academic Integrity: Students enrolled all UT Arlington courses are expected to adhere to the UT Arlington Honor Code:

I pledge, on my honor, to uphold UT Arlington's tradition of academic integrity, a tradition that values hard work and honest effort in the pursuit of academic excellence.

I promise that I will submit only work that I personally create or contribute to group collaborations, and I will appropriately reference any work from other sources. I will follow the highest standards of integrity and uphold the spirit of the Honor Code.

UT Arlington faculty members may employ the Honor Code in their courses by having students acknowledge the honor code as part of an examination or requiring students to incorporate the honor code into any work submitted. Per UT System *Regents' Rule* 50101, §2.2, suspected violations of university's standards for academic integrity (including the Honor Code) will be referred to the Office of Student Conduct. Violators will be disciplined in accordance with University policy, which may result in the student's suspension or expulsion from the University. Additional information is available at <https://www.uta.edu/conduct/>.

Electronic Communication: UT Arlington has adopted MavMail as its official means to communicate with students about important deadlines and events, as well as to transact university-related business regarding financial aid, tuition, grades, graduation, etc. All students are assigned a MavMail account and are responsible for checking the inbox regularly. There is no additional charge to students for using this account, which remains active even after graduation. Information about activating and using MavMail is available at <http://www.uta.edu/oit/cs/email/mavmail.php>.

Campus Carry: Effective August 1, 2016, the Campus Carry law (Senate Bill 11) allows those licensed individuals to carry a concealed handgun in buildings on public university campuses, except in locations the University establishes as prohibited. Under the new law, openly carrying handguns is not allowed on college campuses. For more information, visit <http://www.uta.edu/news/info/campus-carry/>

Student Feedback Survey: At the end of each term, students enrolled in face-to-face and online classes categorized as "lecture," "seminar," or "laboratory" are directed to complete an online Student Feedback Survey (SFS). Instructions on how to access the SFS for this course will be sent directly to each student through MavMail approximately 10 days before the end of the term. Each student's feedback via the SFS database is aggregated with that of other students enrolled in the course. Students' anonymity will be protected to the extent that the law allows. UT Arlington's effort to solicit, gather, tabulate, and publish student feedback is required by state law and aggregate results are posted online. Data from SFS is also used for faculty and program evaluations. For more information, visit <http://www.uta.edu/sfs>.

Final Review Week: for semester-long courses, a period of five class days prior to the first day of final examinations in the long sessions shall be designated as Final Review Week. The purpose of this week is to allow students sufficient time to prepare for final examinations. During this week, there shall be no scheduled activities such as required field trips or performances; and no instructor shall assign any themes, research problems or exercises of similar scope that have a completion date during or following this week *unless specified in the class syllabus*. During Final Review Week, an instructor shall not give any examinations constituting 10% or more of the final grade, except makeup tests and laboratory examinations. In addition, no instructor shall give any portion of the final examination during Final Review Week. During this week, classes are held as scheduled. In addition, instructors are not required to limit content to topics that have been previously covered; they may introduce new concepts as appropriate.

Emergency Exit Procedures: Should we experience an emergency event that requires us to vacate the building, students should exit the room and move toward the nearest exit. When exiting the building during an emergency, one should never take an elevator but should use the stairwells. Faculty members and instructional staff will assist students in selecting the safest route for evacuation and will make arrangements to assist individuals with disabilities.

Students are encouraged to subscribe to the MavAlert system that will send information in case of an emergency to their cell phones or email accounts. Anyone can subscribe at <https://mavalert.uta.edu/> or <https://mavalert.uta.edu/register.php>

Student Support Services: UT Arlington provides a variety of resources and programs designed to help students develop academic skills, deal with personal situations, and better understand concepts and information related to their courses. Resources include [tutoring](#), [major-based learning centers](#),

developmental education, [advising and mentoring](#), personal counseling, and [federally funded programs](#). For individualized referrals, students may visit the reception desk at University College (Ransom Hall), call the Maverick Resource Hotline at 817-272-6107, send a message to resources@uta.edu, or view the information at <http://www.uta.edu/universitycollege/resources/index.php>.

The IDEAS Center (2nd Floor of Central Library) offers **free** tutoring to all students with a focus on transfer students, sophomores, veterans and others undergoing a transition to UT Arlington. To schedule an appointment with a peer tutor or mentor email IDEAS@uta.edu or call (817) 272-6593.

The English Writing Center (411LIBR): The Writing Center Offers free tutoring in 20-, 40-, or 60-minute face-to-face and online sessions to all UTA students on any phase of their UTA coursework. Our hours are 9 am to 8 pm Mon.-Thurs., 9 am-3 pm Fri. and Noon-6 pm Sat. and Sun. Register and make appointments online at <http://uta.mywconline.com>. Classroom Visits, workshops, and specialized services for graduate students are also available. Please see www.uta.edu/owl for detailed information on all our programs and services.

The Library's 2nd floor Academic Plaza offers students a central hub of support services, including IDEAS Center, University Advising Services, Transfer UTA and various college/school advising hours. Services are available during the library's hours of operation. <http://library.uta.edu/academic-plaza>

Resources for Students

Academic Help

- Library Home Page <http://library.uta.edu>
- Academic Plaza Consultation Services library.uta.edu/academic-plaza
- Ask Us ask.uta.edu/
- Library Tutorials library.uta.edu/how-to
- Subject and Course Research Guides libguides.uta.edu
- Subject Librarians library.uta.edu/subject-librarians

Resources

- A to Z List of Library Databases libguides.uta.edu/az.php
- Course Reserves pulse.uta.edu/vwebv/enterCourseReserve.do
- FabLab fablab.uta.edu/
- Special Collections library.uta.edu/special-collections
- Study Room Reservations openroom.uta.edu/

Teaching & Learning Services for Faculty

- Copyright Consultation library-sc@listserv.uta.edu
- Course Research Guide Development, Andy Herzog amherzog@uta.edu
- Data Visualization Instruction, Peace Ossom-Williamson peace@uta.edu
- Digital Humanities Instruction, Rafia Mirza rafia@uta.edu
- Graduate Student Research Skills Instruction, Andy Herzog amherzog@uta.edu
- Project or Problem-Based Instruction, Gretchen Trkay gtrkay@uta.edu
- Undergraduate Research Skills Instruction, Gretchen Trkay gtrkay@uta.edu

Emergency Phone Numbers: In case of an on-campus emergency, call the UT Arlington Police Department at **817-272-3003** (non-campus phone), **2-3003** (campus phone). You may also dial 911. Non-emergency number 817-272-3381

INSY 5378 Spring 2017 Course Schedule (Tentative)

Days	Topics	Readings / Assignments
1/17 1/19	Introduction Data Science Showcases	Syllabus Lecture Note
1/24 1/26	Python Data Types Imperative Programming	LP Ch. 2 / HW#1 Open LP Ch. 3
1/31 2/2	Text Data, File I/O Execution Control	LP Ch. 4 LP Ch. 5 / HW#1 Due (2/3)
2/7 2/9	Containers and Randomness Namespaces	LP Ch. 6 / HW#2 Open LP Ch. 7
2/14 2/16	Object-Oriented Programming Review	LP Ch. 8 HW#2 Due (2/17)
2/21 2/23	Exam 1 Project Brainstorming	LP Chapters 2-8 Lecture Note
2/28 3/2	Data Collection (Web Scraping, BeautifulSoup, Twitter API, etc.)	JG Ch. 9 / Project#1 Open Kick-off Report Due (3/3)
3/7 3/9	Text Mining (NLTK, Gensim, Word Cloud, Sentiment Analysis, Topic Modeling)	JG Ch. 20 Lecture Note
3/14, 3/16	Spring Break – No classes	
3/21, 3/23	Machine Learning Concepts, Data Exploration, Dimensionality Reduction (PCA, MDS)	JG Ch. 10-11 Project#1 Due (3/24)
3/28, 3/30	Classification (KNN, Naïve Bayes, SVM)	JG Ch. 12-13 / Project#2 Open
4/4, 4/6	Regression (Linear, Multiple, Logistic)	JG Ch. 14-16
4/11 4/13	Review Exam 2	JG Ch. 9-16, 20
4/18 4/20	Deep Learning, Image Analysis (TensorFlow, OCR) Clustering (K-Means, Hierarchical)	JG Ch. 18 JG Ch. 19
4/25	Network Analysis (Networkx)	JG Ch. 21 / Project#2 Due (4/26)
4/27, 5/2, 5/4	Final Project Presentations (10 minutes for each group)	Inter-Team Peer Evaluation Report Attendance will be taken
5/9	Submit Final Project Report and Intra-Team Peer Evaluation Report by 11:59 PM	

NOTE: As the instructor for this course, I reserve the right to adjust this schedule in any way that serves the educational needs of the students enrolled in this course. Students are responsible to be aware of changes announced in class and/or via Blackboard. – Gene Moo Lee

Kindly check the Academic Calendar for important dates.