

# INSY 5378: Data Science: A Programming Approach

## Fall 2016

**Instructor:** Gene Moo Lee, Ph.D.

**Office Number:** COBA (Business Building) 516

**Office Telephone Number:** 817-272-3084

**Email Address:** [gene.lee@uta.edu](mailto:gene.lee@uta.edu)

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

**Office Hours:** TR 4:30~5:00 PM, TR 7:00~7:30 PM or by appointment

**Section Information:** INSY 5378, Section 001, Class number 89481

**Time and Place of Class Meetings:** TR 5:30~6:50 PM, COBA 239

**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:

- 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).
- Accessing data from files (e.g., text, CSV, JSON, etc.) and databases (MySQL, MongoDB).
- Text Mining in Python/Scala, including topic modeling and sentiment analysis
- Basics of Social Network Analysis using Networkx
- Machine learning algorithms using Scikit-learn
- 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
- Supplementary Book (for beginning programmers):** *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 with IDE (preferably Anaconda from Continuum Analytics; it can be downloaded at <https://www.continuum.io/downloads>); other software packages as needed

### 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** may be given and quiz scores 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:** 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 other portions (Exam and Projects) will replace the missed exam score. You can only use this excuse for one exam. 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. Homework and project must be turned in electronically (via BlackBoard) by the due date and time specified by the instructor. Late submissions will receive a score of 0 (no exceptions).

**Final Project:** An important component of the course is a group project. The class will be divided into groups of three or four (preferably three) for this purpose. Each group will analyze a dataset, write up a report, and make a presentation to the class. Groups must submit a **kick-off report** (a brief description, problem statement, data, models/algorithms) to the professor by the due date (see Course Schedule) and get his approval before proceeding to work on the project. The project is meant to reinforce concepts taught in the lectures, slides, and/or additional readings. Points will be taken off if you do not attend **all** the presentations. Points for the project will be based on the following criteria:

1. Demo of the project (i.e., a walkthrough)
2. Depth of analysis
3. Communication – how well you present your story
4. Peer evaluations (therefore, it is possible for members of the same group to receive different scores)
5. A project report that has the following:
  - a) The business / research question that you are addressing
  - b) The implications of the project
  - c) Evaluation of models (why was a particular model chosen?)
  - d) Lesson learned

- ***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://www.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](http://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](http://www.uta.edu/disability).

**Counseling and Psychological Services, (CAPS)** [www.uta.edu/caps/](http://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](http://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](http://www.uta.edu/titleIX) or contact Ms. Jean Hood, Vice President and Title IX Coordinator at (817) 272-7091 or [jmhood@uta.edu](mailto: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](mailto:resources@uta.edu), or view the information at <http://www.uta.edu/universitycollege/resources/index.php>.

**The IDEAS Center** (2<sup>nd</sup> 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](mailto: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](http://www.uta.edu/owl) for detailed information on all our programs and services.

The Library's 2<sup>nd</sup> 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](http://library.uta.edu/academic-plaza)
- Ask Us [ask.uta.edu/](http://ask.uta.edu/)
- Library Tutorials [library.uta.edu/how-to](http://library.uta.edu/how-to)
- Subject and Course Research Guides [libguides.uta.edu](http://libguides.uta.edu)
- Subject Librarians [library.uta.edu/subject-librarians](http://library.uta.edu/subject-librarians)

### Resources

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

### Teaching & Learning Services for Faculty

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

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

## INSY 5378 Fall 2016 Course Schedule (Tentative)

Days	Topics	Readings / Assignments
8/25	Introduction	Syllabus
8/30	Big Data and Python	
9/1	Python Data Types	LP Ch. 2
9/6	Imperative Programming	LP Ch. 3
9/8	Text Data, Files	LP Ch. 4
9/13	Execution Control	LP Ch. 5
9/15	Containers and Randomness	LP Ch. 6 / Homework#1
9/20	Namespaces	LP Ch. 7
9/22	Object-Oriented Programming	LP Ch. 8
9/27	Review	Homework#2
<b>9/29</b>	<b>Exam 1</b>	<b>LP Chapters 2-8</b>
10/4	Project Brainstorming	
10/6	Data Collection / Data Preprocessing	JG Ch. 9-10
10/11	Machine Learning / K-Nearest Neighbors	JG Ch. 11-12
10/13	Unsupervised Learning / Clustering	JG Ch. 19
10/18	Text Mining	JG Ch. 20
10/20	Social Network Analysis	JG Ch. 21 / Project#1
10/25	Review	<b>Kick-off Report (10/25 11:59 PM)</b>
<b>10/27</b>	<b>Exam 2</b>	<b>JG Chapters 9-11, 19-21</b>
11/1, 11/3	Regression	JG Ch. 13-16
11/8	Decision Trees	JG Ch. 17
11/10	Recommender Systems	JG Ch. 22
11/15	Visualization	JG Ch. 3
11/17	Deep Learning / Neural Network	JG Ch. 18 / Project#2
11/22, 11/24	Thanksgiving Week – No classes	
<b>11/29</b>	<b>Final Project Presentations</b>	
<b>12/1</b>	<b>Final Project Presentations</b>	
<b>12/6</b>	<b>Final Project Presentations</b>	
<b>12/13</b>		<b>Final Project Report (11:59 PM)</b>

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.**