

**University of Texas – Arlington
School of Business**



**INSY 5378 - Data Science: A Programming Approach
Fall 2019**

Instructor: Mahyar Sharif Vaghefi **Section Information:** INSY 5378
Office Number: 508 **Time:** Mondays 7:00 to 9:50 PM
E-mail: mahyar.sharifvaghefi@uta.edu **Place of Class Meetings:** 339
Office Hours: Monday and Wednesday 4:00 – 5:30 PM, Or by appointment

Course Description

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, Supervised and Unsupervised Learning methods, Text Analytics, and Social Network Analysis.

Prerequisite:

INSY 5336 (Python) and INSY 5339 (Data Mining)

Course Objectives

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.) and databases.
- c. Machine learning algorithms using Scikit-learn.
- d. Performing Text Analysis using Python Tools.
- e. Basics of Social Network Analysis

Textbooks



Python Data Science Handbook – Essential Tools for Working with Data by Jake VanderPlas, 1st edition, O’Reilly Media, Inc., 2016. (Main Book)



Python Data Science Essentials by Alberto Boschetti, and Luca Massaron, 3rd edition, Packt Pub, 2018.



Introduction to Machine Learning with Python by Andreas C. Müller, and Sarah Guido, O’Reilly Media, Inc., 2017.

In-Class Behavior and Electronic Devices

PLEASE ARRIVE ON TIME. Students are expected to behave in a professional and respectful manner. This includes listening attentively, i.e. no texting, checking e-mail, web surfing, reading the paper, sleeping, or talking with other students, etc. Students who do not conduct themselves accordingly will be asked to leave.

Course Requirement

Attendance and participation: At the University of Texas at Arlington, taking attendance is not required. Rather, 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. Keep in mind that missing one session of the course not only impacts on your short-term performance but can also impact on comprehension of topics in future sessions.

Software:

- a. All students need to install Anaconda package for Python version 3 on their laptop (you can find it at <https://www.anaconda.com/download>). Having access to this package is essential for doing assignments, projects and taking exams.
- b. In order to have an effective in-class communication, a web/app-based application (named **Socrative**) will be used to conduct class discussions. Students can use either of the following options to connect to the class discussion board: (1) Install the student version of the Socrative app on their phone (2) Use any web browser to connect to [socrative.com](https://www.socrative.com) using their laptops or in-class computers. This application is free, and there is no cost associated with it for students. Additional information will be provided by the instructor.

Examinations: Quizzes and exams are used to reinforce and integrate learning. There are 3 quizzes and 2 exams (Midterm and Final) in this course. Each taken as scheduled within a limited time. See the *Schedule of Topics* for the dates of the examinations. Please check your work-related schedule early in the semester to ensure that no scheduling conflicts arise.

- **Quizzes**
Quizzes are conducted during class sessions within a short time period. Quizzes contain multiple choice, True/False, and short answer questions. Students are allowed to bring their own notes for quizzes (further instructions will be provided by the instructor). Each quiz comprises 5% of the overall grade.
- **Exams**
Midterm and final exams are comprehensive tests (Midterm exam comprises 20% and Final exam comprises 20% of the overall grade). Students are NOT allowed to bring notes during the exams. Further guidelines will be provided before exams.

Exam Policy: Except for documented emergencies, exams and quizzes cannot be made up except by prior arrangement with the instructor. Such a request must be made in person and in writing, and must include relevant supporting documentations. In case of an emergency, please email or call the instructor as soon as possible.

Individual Assignments:

- **Assignments:**
There are five individual assignments in this course, each comprises 5% of the overall grade. Files should be submitted through the Canvas. No late submission will be accepted. Please do NOT wait for the last minute to deliver your assignments and DO plan for possible computer or network problems. Detailed guidelines will be provided See the *Schedule of Topics* for the due date.

Group Assignments:

- **Group Project:**
An important component of the course is the group projects. The purpose of the group projects is to provide hands-on experience on topics discussed in the class. Hands-on experiences lead to deeper understanding of the class materials. Group members should use discussion forums in Canvas to discuss about different aspects of each project. At the end, each group should submit a complete report in the requested format by the instructor. Detailed guidelines and table of contents will be provided for each project. Projects totally comprises 20% of the overall grade. Project files should be submitted in Canvas. No late delivery will be accepted. Please do NOT wait for the last minute to deliver your assignments and DO plan for possible computer or network problems.

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 an additional 9 hours per week of their own time in course-related activities, including reading required materials, completing assignments, contributing to forums and preparing for exams/quizzes.

Project Teams: Each team consists of 4 to 5 students. All students in a team are working on the group projects. Requests for working individually or with a smaller team cannot be accepted. All team members must take part in the group project activities, although responsibilities may be divided so that different members take lead in different activities. But, *no* activity should be done

exclusively by a single person. While the volume of work of each group member on each project component may not be equal, their contribution to the overall project should equal out.

Here are some *suggestions* (none of this is required!) about team work:

- Choose the team leader to serve as a focal point. Discuss individual skills and strengths and decide who will lead which part of work.
- Set the meeting agenda and time limits. Students are encouraged to confer frequently and work together on solving the problem
- Decide how will the team communicate, e.g., email, Twitter, Facebook, blog, ...
- Make every effort to ensure that *all* team members feel comfortable about other members' contributions
- If in the days before a project deadline you are prevented (travel, illness, ...) from attending the team meetings, immediately notify all your team members and try to contribute to the best of your abilities
- If for whatever reason your contribution falls short of that of other team members, negotiate to assume greater work share in the subsequent deliverable(s)
- General policy: If in doubt, communicate! Every time you send an email related to the project, copy the email to all team members. Do not assume that they are not interested in it or they know what you're talking about. Acknowledge any electronic communication specifically targeted to you.

Saying that “nobody asked me to do this or that, or, I did everything that I was asked to do” is an **unacceptable excuse**. Each team member should be *proactive* and not wait passively to be assigned responsibilities. Do *not* ask others what should be done; rather, take initiative and suggest what should be done to make your project successful. Take every opportunity to redistribute and/or rotate the responsibilities, make your personal suggestions be heard! Many times defining the problem and determining what needs to be done is more difficult than actually doing it. Hence, problem defining and task assignment must be contributed to by all team members, rather than by the team leader alone.

If you notice that your team does not function well, or the team leader tries to misuse his or her role, and this could negatively impact your project performance, you should let me know. Complaints about poor team functioning expressed at the end of semester will be ignored.

Extra Credit: There are three types of extra credit activities in this course.

- **Attendance and Participation:** Students will receive extra credit for their attendance and participation in class activities.
- **In-class Group Competition:** There will be a group competition at the end of the semester between groups. Winner of this competition will receive extra credit points.
- **Best Group Project:** There will be extra points for the group(s) that submitted the best group projects.

Evaluation and Grading Scheme**Evaluation procedure:**

Quizzes:	
• Quiz #1	5%
• Quiz #2	5%
• Quiz #3	5%
Exams:	
• Mid Term Exam	20%
• Final Exam	20%
Individual Assignments:	
• Assignments	25% (5×5%)
Group Assignments:	
• Projects and Presentations	20% (2×10%)
Extra:	
• Attendance and Participation	2%
• In-class Group Competition	2%
• Best Group Project	2%

Total:	106%

Grading scale:

A	>90%
B	<90%-80%
C	<80%-70%
D	<70%-60%
F	<60%

University 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.

<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.

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>.

Active Shooter: Stop. Think. Protect Yourself. You Have Choices.

The safety and security of our campus is the responsibility of everyone in our community. Each of us has an obligation to be prepared to appropriately respond to threats to our campus, such as an active aggressor.

This graphic provides each member of the UTA community with information and options for responding to an active threat. These options are not chronological, but are designed to address dynamic situations. Assess the situation (your location, the location of the threat, type of threat, etc.), identify and weigh your options, develop a plan of action and commit to it.

YOUR OPTIONS TO AN ACTIVE THREAT

You Have Choices!

A V O I D	<ul style="list-style-type: none"> • AVOID the situation. <u>Stay away</u> from the area and campus. • If you can safely leave the area, RUN. • Get others to leave the area, if possible. • Prevent others from entering the area. 	<ul style="list-style-type: none"> • Know your exit and escape options. • If in a parking lot, get to your car and leave. • If in an unaffected area, stay where you are. • When you are safe, call UTA PD at 817.272.3003 or 911 with information you have.
D E N Y	If you can't leave the area safely, DENY or slow entry to the intruder:	
	<ul style="list-style-type: none"> • Lock/barricade doors with heavy items. • Turn off lights/projectors/equipment. • Close blinds and block windows. • Stay away from doors and windows. 	<ul style="list-style-type: none"> • Silence phones and remain quiet. <u>Don't let your phone give you away.</u> • HIDE and take cover to protect yourself. • Be prepared to run or defend yourself.
D E F E N D	If you can't AVOID or DENY entry to the intruder, DEFEND your location:	
	<ul style="list-style-type: none"> • As a last resort, <u>FIGHT for your life</u>. • Use physical force and any weapons available - fire extinguishers, books, chairs, belts, umbrellas, pens/scissors, hot coffee/drinks, trash cans, etc. 	<ul style="list-style-type: none"> • Use the element of surprise. • Work together as a team. Develop a plan. Commit to your actions. Your life depends on it. • Be aggressive, loud, and determined in your actions.

Follow ALL instructions.

For more information, go to: police.uta.edu/activeshooter



Emergency: 817.272.3003
 Non-Emergency: 817.272.3381
police.uta.edu

Additional information for active threat and other emergency situations can be found through the links below:

police.uta.edu/activeshooter

police.uta.edu/em

Schedule of Topics

INSY 5378 - Data Science: A Programming Approach

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 Canvas. *Mahyar Sharif Vaghefi*

Week	Date	Topics	Notes
1	Aug 26	Introduction to Data Science. Overview of the Syllabus and Course Requirements. Python Programming Overview	
	Sep 2	Labor Day	
2	Sep 9	A Glance at Essential Packages: Introduction to Numpy	
3	Sep 16	A Glance at Essential Packages: Introduction to Pandas	Dis: Assignment #1
4	Sep 23	Data Preprocessing and Visualization	Quiz #1
5	Sep 30	Supervised Learning (Part I)	Due: Assignment #1 Dis: Assignment #2
6	Oct 7	Supervised Learning (Part II)	Dis: Project #1
7	Oct 14	Supervised Learning (Part III)	Quiz #2 Due: Assignment #2
8	Oct 21	Unsupervised Learning	Dis: Assignment #3
9	Oct 28	Midterm Exam	The exam is on: 1. Python Overview 2. Numpy and Pandas 3. Preprocessing 4. Visualization 5. Supervised Learning
10	Nov 4	Text Analytics (Part I)	Due: Assignment #3 Due: Project #1 Dis: Project #2
11	Nov 11	Text Analytics (Part II)	Dis: Assignment #4
12	Nov 18	Social Network Analysis	Quiz #3 Dis: Assignment #5
13	Nov 25	Introduction to Orange	Due: Assignment #4
14	Dec 2	Project Presentation	Due: Project #2 Due: Assignment #5
15	Dec 9	Final Exam	The exam is on: 1. Supervised Learning 2. Unsupervised Learning 3. Text Analytics 4. Social Network Analysis