

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



**INSY 5378 - Data Science: A Programming Approach  
Spring 2020**

<b>Instructor:</b> Mahyar Sharif Vaghefi	<b>Section Information:</b> INSY 5378
<b>Office Number:</b> 508	<b>Time:</b> Mondays 7:00 to 9:50 PM
<b>E-mail:</b> <a href="mailto:mahyar.sharifvaghefi@uta.edu">mahyar.sharifvaghefi@uta.edu</a>	<b>Place of Class Meetings:</b> COB 336
<b>Teaching Assistant:</b> Deo Ankit	<b>TA Email:</b> <a href="mailto:ankit.deo@mavs.uta.edu">ankit.deo@mavs.uta.edu</a>

Due to the COVID-19 events, this course switched to online modality as of March 23, 2020. As a result, the assignments from March 23 to the end of the semester will be adjusted to reflect this change. Please note the changes made in the assignments as indicated in Schedule of Topics Section of the Syllabus. The instructor reserves the right to make additional changes to the assignments or grading schema as needed to complete this course on time, and respond to university policies.

### **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:

- 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.
- Machine learning algorithms using Scikit-learn.
- Performing Text Analysis using Python Tools.
- Basics of Social Network Analysis

## **Textbooks**



*Python Data Science Handbook – Essential Tools for Working with Data* by Jake VanderPlas, 1<sup>st</sup> edition, O'Reilly Media, Inc., 2016. (Main Book)



Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 2<sup>nd</sup> edition, O'Reilly Media, Inc., 2019



*Python Data Science Essentials* by Alberto Boschetti, and Luca Massaron, 3<sup>rd</sup> edition, Packt Pub, 2018.



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

## **Course Requirement**

### **Software:**

- 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.

**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 online within a short time period. Quizzes contain multiple choice, True/False, and short answer questions. Students are allowed to use their own notes for quizzes (further instructions will be provided by the instructor). Each quiz comprises 7% 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 use 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 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 four 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.

- **Project:**

There are two individual projects in this course. It comprises 19% 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.

**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.

## Evaluation and Grading Scheme

### Evaluation procedure:

Quizzes:

- Quiz #1 7%
- Quiz #2 7%
- Quiz #3 7%

Exams:

- Mid Term Exam 20%
- Final Exam 20%

Individual Assignments:

- Assignments 20% (4×5%)
- Individual Project1 7%
- Individual Project2 12%

Extra:

• Extra Credit	4%
-----	
Total:	104%

**Grading scale:**

<b>A</b>	>90%
<b>B</b>	<90%-80%
<b>C</b>	<80%-70%
<b>D</b>	<70%-60%
<b>F</b>	<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](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.

**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>.

**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.

<h1 style="margin: 0;">YOUR OPTIONS TO AN ACTIVE THREAT</h1> <h2 style="margin: 0;">You Have Choices!</h2>	
A V O I D	<ul style="list-style-type: none"> <li><b>AVOID</b> the situation. <u>Stay away</u> from the area and campus.</li> <li>If you can safely leave the area, <b>RUN</b>.</li> <li>Get others to leave the area, if possible.</li> <li>Prevent others from entering the area.</li> </ul> <ul style="list-style-type: none"> <li>Know your exit and escape options.</li> <li>If in a parking lot, get to your car and leave.</li> <li>If in an unaffected area, stay where you are.</li> <li>When you are safe, call UTA PD at 817.272.3003 or 911 with information you have.</li> </ul>
D E N Y	<p>If you can't leave the area safely, <b>DENY</b> or slow entry to the intruder:</p> <ul style="list-style-type: none"> <li>Lock/barricade doors with heavy items.</li> <li>Turn off lights/projectors/equipment.</li> <li>Close blinds and block windows.</li> <li>Stay away from doors and windows.</li> </ul> <ul style="list-style-type: none"> <li>Silence phones and <b>remain quiet</b>. <u>Don't let your phone give you away</u>.</li> <li><b>HIDE</b> and take cover to protect yourself.</li> <li>Be prepared to run or defend yourself.</li> </ul>
D E F E N D	<p>If you can't AVOID or DENY entry to the intruder, <b>DEFEND</b> your location:</p> <ul style="list-style-type: none"> <li>As a last resort, <u>FIGHT for your life</u>.</li> <li>Use physical force and any weapons available - fire extinguishers, books, chairs, belts, umbrellas, pens/scissors, hot coffee/drinks, trash cans, etc.</li> </ul> <ul style="list-style-type: none"> <li>Use the element of surprise.</li> <li>Work together as a team. Develop a plan. Commit to your actions. Your life depends on it.</li> <li>Be aggressive, loud, and determined in your actions.</li> </ul>
<p><b>Follow ALL instructions.</b></p> <p>For more information, go to: <a href="http://police.uta.edu/activeshooter">police.uta.edu/activeshooter</a></p> <div style="display: flex; justify-content: space-between; align-items: center;"> <div style="text-align: center;">  </div> <div style="text-align: right;"> <p>Emergency: 817.272.3003</p> <p>Non-Emergency: 817.272.3381</p> <p>police.uta.edu</p> </div> </div>	

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

- [police.uta.edu/activeshooter](http://police.uta.edu/activeshooter)
- [police.uta.edu/em](http://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	Jan 27	Introduction to Data Science. Overview of the Syllabus and Course Requirements. Python Programming Overview	
2	Feb 3	A Glance at Essential Packages: Introduction to Numpy	
3	Feb 10	A Glance at Essential Packages: Introduction to Pandas	Dis: Assignment #1
4	Feb 17	Data Preparation and Visualization	<b>Quiz #1</b> Dis: Individual Project
5	Feb 24	Supervised Learning (Part I)	<b>Due: Assignment #1</b> Dis: Assignment #2
6	Mar 2	Supervised Learning (Part II)	
<b>Spring Break</b>			
<b>Spring Break</b>			
7	Mar 23	Supervised Learning (Part III)	<b>Quiz #2</b> <b>Due: Assignment #2</b>
8	Mar 30	Supervised Learning (Part III) + Unsupervised Learning	Dis: Assignment #3 Dis: Individual Project 2
9	<b>Apr 6</b>	<b>Midterm Exam</b>	The exam is on: 1. Python Overview 2. Numpy and Pandas 3. Preprocessing 4. Visualization 5. Dimension Reduction 6. Supervised Learning
10	Apr 13	Text Analytics (Part I)	<b>Due: Assignment #3</b> Dis: Assignment #4
11	Apr 20	Text Analytics (Part II)	

12	Apr 27	Social Network Analysis	<b>Quiz #3</b> <b>Due: Individual Project 1</b>
13	Apr 27	Advanced Topics	<b>Due: Individual Project 2</b>
15	<b>May 11</b>	<b>Final Exam</b>	The exam is on: <ol style="list-style-type: none"><li>1. Supervised Learning</li><li>2. Unsupervised Learning</li><li>3. Text Analytics</li><li>4. Advanced Topics</li><li>5. Social Network Analysis</li></ol>