**46INSY 5378-001: Data Science: A Programming Approach**

**Summer 2022**

**Instructor Information**

**Instructor:** Sridhar Nerur

**Office:** COBA 518

**Email Address:** snerur@uta.edu

**Office Hours: MW 4:30 to 5:20 p.m.**

**Course Information**

**Section Information:** INSY 5378-001

**Time and Place of Class Meetings:** Mondays and Wednesdays 5:30 – 7:50 p.m. PKH 103 (**NOTE: THIS IS AN IN-PERSON CLASS)**

**Catalog 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. This course provides an in-depth understanding of data preprocessing/feature engineering as well as machine learning concepts and algorithms using Python. Students will receive hands-on training on supervised learning algorithms such as KNN, Naïve Bayes, Linear and Logistic Regression, Support Vector Machines, Decision Trees and Ensembles, and Artificial Neural Networks (ANNs). The course will also cover the foundations of Natural Language Processing (NLP), Social Network Analysis, and unsupervised learning algorithms such as K-Means, Hierarchical Clustering, t-SNE, and DBSCAN.

**Prerequisite**: [INSY 5336](http://catalog.uta.edu/search/?P=INSY%205336) (Python) and [INSY 5339](http://catalog.uta.edu/search/?P=INSY%205339) (Data Mining)

**NOTE**: This is a **hands-on** course that requires a good grasp of Python. If you wish to brush up on your Python, you may download easy-to-use Jupyter Notebooks covering various aspects of Python from<https://github.com/snerur/Introduction-to-Python>. I think you will find them useful.

The course also assumes that you have a good understanding of linear algebra, calculus (particularly derivatives), and statistics & probability. The following blog provides a list of useful resources for reviewing concepts from these topics.

<https://towardsdatascience.com/mathematics-for-data-science-e53939ee8306>

**Learning 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), and understand how these are used in machine learning.

b. Market-Basket Analysis

c. Text Analytics in Python, including topic modeling and sentiment analysis

d. Supervised and unsupervised (Machine) learning algorithms using Scikit-learn.

**Required Textbooks and Other Course Materials:**

* *Python Machine Learning By Example: Building intelligent systems using Python, Tensorflow 2, PyTorch, and scikit-learn* by Yuxi (Hayden) Liu, Third Edition, Packt Publishing, 2020, ISBN: 9781800209718

**Other Materials/Resources/Recommended Readings:**

1. <https://developers.google.com/machine-learning/crash-course>
2. <http://www.codecademy.com/en/tracks/python>
3. <https://developers.google.com/edu/python/>
4. Also check out [www.coursera.org](http://www.coursera.org) and [www.udacity.com](http://www.udacity.com) for introductory Python courses
5. <http://www.astro.washington.edu/users/vanderplas/Astr599/notebooks/17_SklearnIntro>
6. *Data Science from Scratch* by Joel Grus, O’Reilly Media, Inc., 2015.

**Software Requirements:**

**Python:** Given below is a list of IDEs:

Anaconda (<http://continuum.io/downloads>) – my first choice

Enthought Canopy (<https://store.enthought.com/downloads/>)

PyCharm (<https://www.jetbrains.com/pycharm/download/>)

IDLE ([www.python.org](http://www.python.org))

**Other Materials/Resources/Readings:**

Notes/slides will be posted in Canvas

**Description of major assignments and examinations:**

The distribution of points will be as follows:

Homework Assignments 15%

Exam 1 20%

MidTerm Exam 25%

Project 15%

Final Exam 25%

**Grading:** The following criteria will be used to assess your grade:

A (>=90%), B (>=80%), C (>=70%), D (>=60%), F (<60%)

**THERE WILL BE NO ROUNDING. TO MAKE AN “A” YOU NEED TO HAVE AN AVERAGE OF 90 OR MORE (89.999 will be a”B”).**

**Final Project:** The final project will be done in groups of no more than 3 students. You need to start thinking about the project as early as possible. The easiest thing to do would be to pick an ongoing project – not a completed one - from Kaggle. Each group will get about 10 minutes to present its project at the end of the semester. Details will be provided in class.

**Attendance:** At the University of Texas at Arlington, taking attendance is not required but attendance is a critical indicator of 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. 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 must 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 Canvas. 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 will be **no make-up exams**. Under extenuating circumstances (e.g., medical emergency, family emergency, etc.), other arrangements may be considered. You will have to take your exams in class. Details will be provided in class.

**Homework Assignments**: You will have 4 or 5 homework exercises. Homework must be uploaded to Canvas by the due date and time specified by the instructor. Late submissions will receive a score of 0. **PLEASE DO NOT EMAIL YOUR HOMEWORKS/ASSIGNMENTS TO ME.**

**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 8 to 16 hours per week of their own time in course-related activities, including reading required materials, completing assignments & projects, and preparing for exams/quizzes.

**Tentative Course Schedule**

|  |  |  |
| --- | --- | --- |
| **Day** | **Assignments and Topics** | **Readings** |
| Week 1  June 6 & 8 | Syllabus; Introduction; Overview of Numpy and Pandas; | Syllabus; Slides; Notebooks |
| Week 2  June 13 & 15 | Preprocessing/Feature Engineering | Slides/Notebooks/Chapter 1 |
| Week 3  June 20 & 22 | Supervised Learning; Bias vs Variance  **Exam 1 on June 22nd (EVERYTHING COVERED IN THE FIRST TWO WEEKS)** | Slides/Notebooks; Chapters 2 to 5 and 8 |
| Week 4  June 27 & June 29 | Supervised Learning (Naïve Bayes, SVM, Linear Models, Trees, Ensembles, Neural Networks) | Chapter 2-5, Chapter 8/Slides/Notebooks |
| Week 5  July 6 | Supervised Learning | Chapters 2-5, Chapter 8 / slides / notebooks |
| Week 6  July 11 and 13 | Unsupervised Learning  **EXAM 2 (14th) over everything that was covered until & including week 5** | slides / notebooks |
| Week 7  July 18 & 20 | Text Analytics/NLP | Chapters 9 & 10; slides/notebooks |
| Week 8  July 25 & 27 | Project Presentations; Final Exam Review | Chapters 3,4, & 5 |
| FINAL EXAM | **COMPREHENSIVE FINAL EXAM** – Date will be provided |  |

NOTE: *This is a tentative schedule and may be changed to serve the educational needs of the students enrolled in the course. It is your responsibility to be aware of changes announced in class and/or via Canvas. – Sridhar Nerur*

**Kindly check the Academic Calendar for important dates**.

**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. Policy details may be obtained from:

<http://catalog.uta.edu/academicregulations/grades/#graduatetext>. <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:** UTArlington 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)** <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).

Only those students who have officially documented a need for an accommodation will have their request honored. 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) or by calling the Office for Students with Disabilities at (817) 272-3364.

**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://www.uta.edu/hr/eos/index.php).

**Title IX:** 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. Michelle Willbanks, Title IX Coordinator at (817) 272-4585 or [titleix@uta.edu](mailto:titleix@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/>. Faculty are encouraged to discuss plagiarism and share the following library tutorials <http://libguides.uta.edu/copyright/plagiarism> and <http://library.uta.edu/plagiarism/>

**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 classes categorized as “lecture,” “seminar,” or “laboratory” shall be 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 enters the SFS database anonymously and is aggregated with that of other students enrolled in the course. UT Arlington’s effort to solicit, gather, tabulate, and publish student feedback is required by state law; students are strongly urged to participate. For more information, visit <http://www.uta.edu/sfs>.

**Final Review Week:** 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.

**Student Success Programs:** 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 by appointment](http://www.uta.edu/studentsuccess/learning-center/utsi/tutoring/index.php), [drop-in tutoring](https://www.uta.edu/ideas/services/index.php), [etutoring](https://www.etutoring.org/login.cfm?institutionid=388&returnPage), [supplemental instruction](http://www.uta.edu/studentsuccess/learning-center/utsi/supplemental-instruction/index.php), [mentoring](https://www.uta.edu/ideas/services/mentoring/index.php) (time management, study skills, etc.), [success coaching](http://www.uta.edu/studentsuccess/success-programs/success-coaching.php), [TRIO Student Support Services](http://www.uta.edu/studentsuccess/learning-center/sss/index.php), and [student success workshops](http://www.uta.edu/studentsuccess/success-programs/success-series-workshops.php). For additional information, please email [resources@uta.edu](mailto:resources@uta.edu), or view the [Maverick Resources](http://www.uta.edu/studentsuccess/success-programs/programs/resource-hotline.php) website.

### Research or General Library Help

Ask for Help

* [Academic Plaza Consultation Services](https://library.uta.edu/academic-plaza)(library.uta.edu/academic-plaza)
* [Ask Us](http://ask.uta.edu/)([ask.uta.edu/](http://ask.uta.edu/))
* [Research Coaches](https://library.uta.edu/subject-librarians) (http://libguides.uta.edu/researchcoach)

Resources

* [Library Tutorials](https://library.uta.edu/how-to) ([library.uta.edu/how-to](http://library.uta.edu/how-to))
* [Subject and Course Research Guides](https://libguides.uta.edu/) ([libguides.uta.edu](http://libguides.uta.edu/))
* [Librarians by Subject](https://library.uta.edu/subject-librarians) (library.uta.edu/subject-librarians)
* [A to Z List of Library Databases](https://libguides.uta.edu/az.php) (libguides.uta.edu/az.php)
* [Course Reserves](https://uta.summon.serialssolutions.com/#!/course_reserves)(https://uta.summon.serialssolutions.com/#!/course\_reserves)
* [Study Room Reservations](https://openroom.uta.edu/)(openroom.uta.edu/)

You may also find the following information useful:

Library Home Page <http://www.uta.edu/library>

Subject Guides <http://libguides.uta.edu>

Subject Librarians <http://www.uta.edu/library/help/subject-librarians.php>

Course Reserves <http://pulse.uta.edu/vwebv/enterCourseReserve.do>

Library Tutorials <http://www.uta.edu/library/help/tutorials.php>

Connecting from Off- Campus <http://libguides.uta.edu/offcampus>

Ask A Librarian [http://ask.uta.edu](http://ask.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 817-272-338