TEXAS WOMAN'S UNIVERSITY

Course Syllabus

Course Information

Course Number and Title	CSCI 3603 Foundations of Data Science
College and Department	School of the Sciences, Division of Computer Science
Term	Fall
Year	2024
Format	Online
Class day(s) and time	Thursday, 5:00 PM to 6:00 PM
Class Location	Online
Weeks in length	16
Class meetings per term	15
Hours per class session	3
Credit hours awarded	3

Instructors Information

Instructor	Islam Akef Ebeid
Contact Number	(940) 898-2165
E-mail	iebeid@twu.edu
Office Hours	Mondays, Tuesdays, Friday 3:00 - 5:00 PM
Mode	Drop by, appointment outside these hours
Office	MCL 412
Zoom	https://twu-edu.zoom.us/j/8362252819

Course Description

This course introduces data science concepts, including programming, data modeling, data management, data manipulation, data analysis techniques, decision-making from rich data sets, information visualization, data mining, and machine learning. The primary central skill that will be learned from this course is how to extract knowledge, insights, and predictions from data. Extracting knowledge from data requires an integrated skill set spanning statistics, machine learning, databases, algorithms, and other branches of computer science and mathematics. Concepts, techniques, and tools needed for diverse facets of data science practice, including data collection and integration, data cleaning, exploratory data analysis, predictive and other types of modeling, visualization and animation, evaluation, interpretation, and effective communication, will all be covered with various depths in this course.
Prerequisites: MATH 1713, CSCI 3423, CSCI 5103

Course Objectives and Learning Outcomes

Today, many important decisions made by individuals, organizations, institutions, governments, and society are data-driven. Critical in the decision-making process today is knowledge of data analytics, business intelligence, and data science. Hence, this course aims to teach students the skills required in today's dynamic Data Science field. The skills necessary to teach this course are grouped into four stages modeled after how a data analysis process undergoes: data preparation, exploration, modeling, experimentation, evaluation, and communication. First, students will be introduced to different types of data: structured and unstructured. In other words, they will be introduced to data science from the perspective of data. That will include data types such as text, images, networks, and databases and data modeling techniques such as relational and graph data models. Second, students will study different types of statistical analysis and exploration. Then, they will be introduced to probability and machine-learning models. This will include parametric and nonparametric methods and other data analysis tasks such as classification, clustering, and curve fitting. Third, students will learn how to design experiments to evaluate the validity of their models. And finally, students will learn how to interpret and explain their models appropriately to extract knowledge, insights, and predictions.

Python was chosen as a programming language in this course because it is highly demanded in the job market. Moreover, Python provides a flexible platform for functional, procedural, and object-oriented programming paradigms. The syntax is easy and helps the student understand basic programming concepts in the least amount of time. It is essential to understand that Python is just a tool to teach students how to think like a data scientist. The core skill in this course is developing proper models to extract knowledge, insights, and predictions from rich and complex data sets. Due to the nature of data science, the topics tend to get complicated quickly. Hence, I employ a step-by-step approach where I don't progress with students to the next step until I am sure that the prerequisite skills are mastered.

Core learning outcomes:

- Appropriately assemble datasets for analysis, including obtaining, cleaning, manipulating, and dealing with missing data
- Correctly apply statistical and mathematical analysis techniques on data
- Correctly analyze datasets using data science techniques
- Critically interpret important trends and findings from data visualization results.

Skills:

- Applying concepts in statistics and probability like descriptive, frequentist, Bayesian, parametric, and nonparametric methods
- Successfully leverage a combination of programming and software tools to complete all the steps in a data science project.
- Applying concepts in statistics and probability like descriptive, frequentist, Bayesian, parametric, and nonparametric methods.
- Getting introduced to different types and formats of datasets: text, images, networks, and databases, and how to preprocess and model those data sets.
- Learning Python as the primary programming tool in Data Science.
- Learning to design experiments to evaluate models.
- Understanding how to communicate results to stakeholders.
- Understanding how to model complex data sets using Python.
- Critically interpreting essential trends and findings from data visualization results.
- Accurately assess the data quality in the preprocessing stage and apply data cleaning techniques accordingly.
- The student shall come out of the class with the skills necessary to excel in more advanced graduate-level courses and topics in Data Science.

Course Format
During the semester, the instructor will post recorded lectures at the beginning of each week. The recorded lecture will be around 2 hours. In addition, the instructor will meet with the students online for 1 hour. The class meeting is optional, and the students do not have to attend except if they have questions. Each week, there will be a multiple-choice questions quiz to assess understanding of the material. The quiz will be due the following week. In addition, there will be weekly assignments where students will be asked to solve 2 to 4 challenges. The challenges vary by topic, from data preprocessing skills to statistical modeling to developing a machine learning model.

Since this class is bracketed, graduate students will be expected to work on a final project individually or in groups. Each individual or group will submit a final project proposal by the fourth week. By the tenth week, graduate students participating in the final project will be expected to provide a progress report on their projects. By the fifteenth week, the students will be expected to have finished the project and submit a final paper laying out their motivation, method, experiments, and result discussion. The paper shall be at least 5 to 6 pages in an appropriate academic format. In group projects, tasks and roles shall be explained clearly in the proposal, progress, and final report.

This course will have a final exam in 2 parts: a multiple-choice question and a problem-solving part.

Course Material
The required textbooks are:

Title	Code	Author	Link
Introduction to Linear Algebra	[STRANG]	Gilbert Strang	http://students.aiu.edu/submissions/profiles/resources/onlineBook/Y5B7M4_Introduction_to_Linear_Algebra-_Fourth_Edition.pdf
Think Python	[DOWNEY]	Allen Downey	https://greenteapress.com/wp/think-python-2e/
Python Data Science Handbook	[PDSH]	Jake VanderPlas	https://jakevdp.github.io/PythonDataScienceHandbook/
Database Design 2nd Edition	[DDE]	Adrienne Watts	https://opentextbc.ca/dbdesign01/
Pattern Recognition and Machine Learning	[BISHOP]	Christopher Bishop	https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf
Deep Learning	[DL]	Ian Goodfellow	http://imlab.postech.ac.kr/dkim/class/csed514_2019s/DeepLearningBook.pdf

The books are free as offered by the publisher or the author through the links they provided, and I posted here. There will also be slides for each class. The slides will be uploaded regularly to Canvas. The assignments, quizzes, and final exams will mainly come from the required textbooks and slides. Assignments will be inspired by multiple sources such as hackerrank.com, leetcode.com, and Kaggle.com. Those online communities offer plenty of solvable activities for students to review and engage with on their own time.

Course flow

Grade Structure

Criteria	Number of Occurrences per Week	Points per Occurrence	Total Points	Percentage of Total
Reading Quizzes	10/16	20	200	20%
Homework Assignments	10/16	40	400	40%
Final Exam	2/16	100	200	20%
Final Project	3/16	50+50+100	200	20%
Total	16/16	-	1000	100%

Grade Policy

A	90%-100%
B	80%-90%
C	70%-80%
D	50%-70%
F	Below 50%

Please note that you must provide a correct solution and answer to score a full grade on assignments. Incorrect solutions will be considered if sufficient effort is shown and demonstrated. 100% scores will be reserved for extraordinary answers.

Course Plan

Week	Topic	Subtopics	Readings	Activity
1 **August 26**	Introductions What is Data Science?	Curve Fitting Data, Models, Experimentation, and Interpretation	Syllabus	Survey
2 **September 2**	From the Perspective of Data	Data Modeling Data Types Data Structures Data Preprocessing Introduction to Python	Slides Downey Think Python: Chapters 1,2,3 Bishop PRML: Chapter 1; Sections 1.1, 1.2 (Gaussian Distribution only)	Quiz 1 – September 11
3 **September 9**	Python for Data Science	Basic Python Programming Python Notebooks Advanced Libraries for Data Science Algorithms Files and Persistence	Slides Downey Think Python: Chapters 11,12,14	Quiz 2 – September 18 Assignment 1 – September 18
4 **September 16**	Math for Data Science	Calculus Linear Algebra Iteration Conditionals Recursion	Slides Goodfellow Deep Learning: Chapter 2 Strang Calculus: Chapters 2 Downey Think Python: Chapter 5, 7	Quiz 3 – September 25 Assignment 2 – September 25
5 **September 23**	Statistics and Probability for Data Science	Probability Theory Non-parametric Statistics Binary Variables Multinomial Variables Algorithms	Slides Bishop: Chapter 1, Section 1.2 Goodfellow: Chapter 2 Downey: Analysis of Algorithms https://allendowney.github.io/DSIRP/analysis.html Python for Data Science: Pages 33-40	Final Project Proposal – October 3 Quiz 4 – October 3 Assignment 3A – October 3
6 **September 30**	Data Modeling for Data Science	Modeling Data Relational Data Models Database Management Systems Files SQL	Slides Think Python: Chapter 14 Files, 15 Classes Database Design 2nd Edition 1-2	Quiz 5 – October 10 Assignment 3B – October 10
7 **October 7**	Machine Learning for Data Science	Optimization Classification Clustering Prediction	Slides Database Design 2nd Edition 3-5 Goodfellow: Chapter 3	Quiz 6 – October 17 Assignment 5 – October 17

		Inference Curve Fitting		
8 **October 14**	Machine Learning for Data Science	Linear Regression Json and XML formats Analysis of Algorithms Numpy, Scipy and SKLearn	Slides Goodfellow: Chapter 4	Quiz 7 – October 24 Assignment 6 – October 24
9 **October 21**	Machine Learning for Data Science	Principle Component Analysis Matrix Factorization	Slides Bishop: Chapter 1 Goodfellow: Chapters 4 chosen sections will be announced	Quiz 8 – October 27 Assignment 7 – October 27
10 **October 28**	Advanced Machine Learning	Natural Language Processing Large Language Models	Slides Goodfellow: Chapters 5 chosen sections.	Quiz 9 – November 3
11 **November 4**	Advanced Machine Learning	Convolutional Neural Networks Object Detection 3D Vision	Slides Python Data Science Handbook: Chapter 5	Assignment 8 – November 3 Progress Report – November 10
12 **November 11**	Advanced Machine Learning	Network Science Graphical Models Graph Representation Learning	Slides Python Data Science Handbook: Chapter 5	Quiz 10 – November 10
13 **November 18**	Thanksgiving break	Thanksgiving break	Thanksgiving break	Thanksgiving break
14 **November 25**	Model Evaluation in Data Science, Ethics in Data Science	Interpretation Ethics in Data Science Decision Theory Information Theory Validation & Evaluation Experimental Design	Slides	
15 **December 2**	Course Recap	Course Recap	Study Guide	Final Project Due
16 **December 9**	Final Exam	Final Exam	Final Exam	Final Exam

Academic Resources

[redacted]

- If you are having trouble finishing an assignment, seek help.

- A good grade could be achieved if you showed effort and explained your thought process despite having an incorrect result.
- See your errors and mistakes as opportunities to learn more.
- I will provide you with feedback. Sometimes, the input will sound like criticism. Not everyone likes that. Please understand that the feedback comes from my intention to ensure you know the content and build the necessary skills.
- Show effort instead of focusing on getting everything right. I grade based on effort.

[illegible]
Please let me know if you need a computer or a laptop for this class. If you need any assistance with technology, please reach out to the IT Solutions (https://twu.edu/technology/)

[illegible]
Please don't buy textbooks or software that you might need for this class before checking in at the library first. https://twu.edu/library/. If you have any questions about the library or how to find a resource, please get in touch with me or Susan Whitmer (swhitmer@twu.edu)

[illegible]
Minerva's Market is in The Student Union at Hubbard Hall, Room 1203
Social Work Food Pantry is in the Old Main Building, Room 406

[illegible]
If you need help with any issue that is affecting your academic performance, please refer to:
https://twu.edu/student-health-services/mental-health/

or

https://twu.edu/counseling/

If you need immediate help, please go directly to Jones Hall Room 269 (M-F 8 AM to 6 PM) or call the Crisis Line: (940) 898-4357

[illegible]
If you have difficulty communicating in written English language, please let the instructor know, and please refer to the following:

https://twu.edu/write-site/

[illegible]
If you would like additional help for the class or any other classes, please notify the instructor and refer to:

https://catalog.twu.edu/graduate/services-available-students/tutoring-centers/

University Policies
For general university policies, please refer to https://web.saumag.edu/academics.

Disability Access Policy Statement
Texas Woman's University strives to make all learning experiences accessible. If you anticipate or experience academic barriers based on your disability (e.g., mental health conditions, learning disabilities, chronic medical conditions, etc.), please register with Disability Services for Students (DSS) to establish reasonable academic accommodations. After registration with DSS, please contact me to discuss how to implement your accommodation.

DSS contact information: DSS website (https://twu.edu/disability-services/); dss@twu.edu; 940-898-3835; CFO Ste. 106.

If you have any questions regarding disability, please reach out to me or Nadaya Cross (ncross1@twu.edu)

Title IX: Sexual Violence Education

TWU is committed to providing a safe learning environment free of all forms of sexual misconduct, including sexual harassment, sexual assault, domestic violence, dating violence, and stalking. Federal laws (Title IX and the Violence Against Women Act) and TWU policies prohibit discrimination based on sex and therefore prohibit sexual misconduct. As students, if you or someone you know is experiencing sexual harassment, relationship violence, stalking, or sexual assault, there are campus resources available to provide support and assistance. Alleged sexual misconduct can be non-confidentially reported to the Title IX Coordinator at the Report an Incident website (https://twu.edu/civility/report-an-incident/) or at (940) 898-2968. Additionally, please be aware that under Title IX of the Education Amendments of 1972, all employees must disclose information about such misconduct to the Title IX Office. Students who wish to speak to a confidential employee who does not have this reporting responsibility can contact TWU Counseling and Psychological Services at (940) 898-3801 for the Denton Campus, (214) 689-6655 for the Dallas Campus, and (713) 794-2059 for the Houston Campus.

Title IX: Pregnant Students

Title IX is a federal law that requires schools that receive federal funds to provide reasonable accommodation to students who are pregnant or have pregnancy-related conditions. This includes pregnancy, pre-natal doctor appointments, childbirth, false pregnancy, miscarriage, termination of pregnancy, or recovery from any of these conditions. Students needing academic accommodations due to pregnancy-related conditions should complete the Pregnancy Accommodation form (https://twu.edu/pregnancy-accommodation-form/) to coordinate educational needs.

Academic Integrity

Honesty in completing assignments is essential to the mission of the University and the development of the personal integrity of students. In submitting graded assignments, students affirm that they have neither given nor received unauthorized assistance and abided by all other provisions of the Academic Integrity Policy and the Student Code of Conduct as found on the TWU website and in the TWU Student Handbook. Cheating, plagiarism, collusion, dual submission of a paper, or other academic dishonesty will not be tolerated. It will result in appropriate sanctions, including failing an assignment, failing the class, being removed from an educational program, or being suspended or expelled. Allegations of academic dishonesty in this course may be reported to the Office of Civility and Community Standards. The specific disciplinary process for academic dishonesty is in the TWU Student Code of Conduct (https://public.powerdms.com/TWU1/documents/1745742) and Academic Integrity Academic Integrity Policy (https://public.powerdms.com/TWU1/documents/1748544). For details on avoiding plagiarism, review the library Tutorial: Avoiding Plagiarism (https://libguides.twu.edu/c.php?g=270163&p=1803990).

To ensure the integrity of the academic process, Texas Woman's University vigorously affirms the importance of academic honesty as defined by the Academic Integrity Policy and the TWU Student Code of Conduct. Therefore, Texas Woman's University faculty members may use Turnitin to compare a student's work with multiple sources to detect and prevent plagiarism. It then reports a similarity percentage and provides links to those specific sources. The tool itself does not determine whether a paper has been plagiarized. Instead, that judgment must be made by the individual faculty member. Some of the required assignments in this course may be checked for plagiarism using Turnitin.com.

Attendance Policy

Consistent attendance is vital to academic success and is expected of all students. Grades are determined by academic performance, and instructors may give students written notice that attendance related to specific classroom activities is required. Absence does not exempt students from academic requirements. Even if documented, excessive absences may result in a student's failing the course. Excused absences are within the purview of the instructor. Students must consult with instructors regarding make-up work.

Departmental Policies

G[illegible] P[illegible]

https://docs.google.com/document/d/1eeTJG916awbljyMG6zIOsuak2U_ozbCN/edit?usp=drive_link

AI U[illegible] P[illegible]

https://docs.google.com/document/d/1rPm6TjS8FRFTLbWD9ERbbBySfXCcHg9j/edit?usp=drive_link

[illegible]
https://docs.google.com/document/d/1n2yJvdRV2BzUb0QezPLeZwJyUJp_eE3Z/edit?usp=drive_link

Instructor Policies

[illegible]
Please respect that this is a respectable higher education institution and behave accordingly. Please keep your communication with your instructor and colleagues positive and constructive.

[illegible]
Late work is not accepted unless there is a compelling reason. If this is the case, please communicate promptly and appropriately.

[illegible]
I rely on Turnitin to gauge the level of plagiarism in your work. This applies to writing and coding. The above 30% similarity is too much, and I consider this plagiarism. Using internet resources is allowed with restrictions that the instructor will mention during the classroom. Yet please don't copy and paste code or answers for any questions.

[illegible]
The instructor will follow the federally and state-recognized holiday schedule by the University, which can be found here:

https://twu.edu/media/documents/registrar/Calendar-at-a-Glance-2022-2023.pdf

However, if you need special accommodation for religious or other types of holidays that you observe, please let the instructor know beforehand.

[illegible]
The instructor is committed to diversity, inclusion, and equality in the classroom and by the university policies, regardless of cultural background, country of origin, religion, race, ethnicity, and sexual orientation. Please let the instructor know how you would like to be addressed. During the first lecture, the instructor will ask the students about their names, pronouns, and other forms of addressing they want to be referred to. Please notify the instructor if that changed or if you were addressed mistakenly.

[illegible]
Generally, email is the best way to reach me. You may drop by my office anytime within the specified hours if you need help. However, you should email first because multiple people might come simultaneously. If that's the case, I will meet with people on a first come, first served basis. I will also have to limit the sessions to 15 minutes if people are waiting. If two or more students come simultaneously, it will be in the order of the last name. Please come prepared to office hours with questions.

[illegible]
In case of conflict between students in the classroom, the instructor will act as a mediator until proper university authorities or public safety are notified.