



Course Syllabus

Course Information

Course Number and Title: CSCI 3603/5573 Foundations of Data Science

College and Department: School of the Sciences, Division of Computer Science

Term: Fall 2023

Format: Online

Class day(s) and time: Thursday, 1:00 PM to 3:00 PM and 6:00 PM to 8:00 PM

Class Location: Online via Zoom <https://twu-edu.zoom.us/j/8362252819>

Weeks in length: 16

Class meetings per term: 15

Hours per class session: 3

Credit hours awarded: 3

Instructor Information

Instructor: Dr. Islam Akef Ebeid

Contact Number: 940 898 2165

E-mail: iebeid@twu.edu

Office Hours: Mondays, Wednesdays, Fridays 3:00 – 5:00 PM

Office: MCL 412

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, evaluation, and communication are the four stages. 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 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.

Here are the core student learning outcomes per bracket for this course:

- 1- CSCI 3603/5573: Appropriately assemble datasets for analysis, including curating, cleaning, integrating, and dealing with missing data.
- 2- CSCI 3603/5573: Successfully leverage a combination of programming and software tools to complete all the steps in a data science project.
- 3- CSCI 3603/5573: Correctly apply statistical and mathematical analysis techniques to data.
- 4- CSCI 3603/5573: Applying concepts in statistics and probability like descriptive, frequentist, Bayesian, parametric, and nonparametric methods.
- 5- CSCI 3603/5573: Getting introduced to different types and formats of datasets: text, images, networks, and databases, and how to preprocess and model those data sets.
- 6- CSCI 3603/5573: Learning Python as the primary programming tool in Data Science.
- 7- CSCI 3603/5573: Learning to design experiments to evaluate models.
- 8- CSCI 3603/5573: Understanding how to communicate results to stakeholders.
- 9- CSCI 3603/5573: The student will come out of this course with an understanding of how to model complex data sets using Python.
- 10- CSCI 3603/5573: Critically interpreting essential trends and findings from data visualization results.
- 11- CSCI 3603/5573: Accurately assess the data quality in the preprocessing stage and apply data cleaning techniques accordingly.
- 12- CSCI 3603: 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.
- 13- CSCI 5573: Digesting more complex issues such as Deep Learning, Bayesian Sampling, and estimation, and analyzing Graph data.
- 14- CSCI 5573: Successfully compose a literature review on scholarly research and case studies 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 1 hour. In addition, the instructor will meet with the students online for 2 hours. The class meeting is optional, and the students do not have to attend except if they have questions. Each week, there will be a 5-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 independently. 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 eight 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:

Introduction to Linear Algebra by Gilbert Strang

[http://students.aiu.edu/submissions/profiles/resources/onlineBook/Y5B7M4 Introduction to Linear Algebra- Fourth Edition.pdf](http://students.aiu.edu/submissions/profiles/resources/onlineBook/Y5B7M4%20Introduction%20to%20Linear%20Algebra-Fourth%20Edition.pdf)

Calculus by Gilbert Strang

<https://ocw.mit.edu/ans7870/resources/Strang/Edited/Calculus/Calculus.pdf>

Think Python by Allen Downey

<https://greenteapress.com/wp/think-python-2e/>

Python Data Science Handbook by Jake VanderPlas

<https://jakevdp.github.io/PythonDataScienceHandbook/>

Pattern Recognition and Machine Learning by Christopher Bishop

<https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>

Network Science by Albert-Laszlo Barabasi

<http://networksciencebook.com/>

Computer Vision: Algorithms and Applications by Richard Szeliski

<https://szeliski.org/Book/>

Deep Learning by Ian Goodfellow

http://imlab.postech.ac.kr/dkim/class/cs514_2019s/DeepLearningBook.pdf

Natural Language Processing with Python by Steven Bird et al.

<https://tjzhifei.github.io/resources/NLTK.pdf>

Additional resources that I might refer you to:

Elements of Data Science

<https://allendowney.github.io/ElementsOfDataScience/README.html>

Think Statistics

<https://greenteapress.com/wp/think-stats-2e/>

Think Bayesian

<https://greenteapress.com/wp/think-bayes/>

Complexity Analysis

<https://greenteapress.com/complexity2/>

Database Design

<https://open.umn.edu/opentextbooks/textbooks/354>

Computational and Inferential Thinking

<https://inferentialthinking.com/chapters/intro>

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.

Grade Structure

Criteria	Number of Occurrences	Points per Occurrence	Total Points	Percentage of Total
Reading Quizzes	13/16	20	260	16-26%
Homework Assignments	11/16	40	440	30-45%
Final Exam	2/16	150	300	10-15%
Final Project (CSCI 5573)	5/16	100	500	0-33%
Total	16/16	-	1000 (CSCI 3603) 1500 (CSCI 5573)	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 28	Introductions What is Data Science?	Curve Fitting Data, Models, Experimentation, and Interpretation	Syllabus	Survey
2 September 4	From the Perspective of Data	Data Modeling Data Types Data Structures Data Preprocessing Introduction to Python	Slides Downey: Chapters 1,2,3 Bishop: Chapter 1; Sections 1.1, 1.2 (Gaussian Distribution only)	Quiz 1 – September 11
3 September 11	Python for Data Science	Basic Python Programming Python Notebooks Advanced Libraries for Data Science Algorithms Files and Persistence	Slides Downey: Chapters 10,11,12,14	Quiz 2 – September 18 Assignment 1 – September 18
4 September 18	Math for Data Science	Calculus Linear Algebra Algorithms Optimization, Recursion, and Iteration	Slides Downey: Appendix B Bishop: Appendix C, D Strang Linear: Chapters 1, 2 and 7 Strang Calculus: Chapters 2, 3 and 4 Goodfellow: Chapter 2	Quiz 3 – September 22 Assignment 2 – September 22
5 September 25	Statistics and Probability for Data Science	Probability Theory Non-parametric Statistics Binary Variables Multinomial Variables	Slides Bishop: Chapter 1, Section 1.2, Chapter 2 Sections 2.1, 2.2, 2-5 Goodfellow: Chapter 3	Final Project Proposal – September 29

				Quiz 4 – September 29 Assignment 3 – September 29
6 October 2	Advanced Statistics and Probability for Data Science	Parametric Statistics Gaussian Distribution Exponential Distributions Bayesian Statistics Sampling Parameter Estimation	Slides Bishop: Chapter 2, Section 2.3, 2.4 Goodfellow: Chapter 3	Quiz 5 – October 6 Assignment 4 – October 6
7 October 9	Machine Learning for Data Science	Classification Clustering Prediction Inference Curve Fitting Linear Regression Non-Linear Regression	Slides Bishop: Chapter 3, 4, Chapter 9, Section 9.1 Goodfellow: Chapter 5	Quiz 6 – October 13 Assignment 5 – October 13
8 October 16	Advanced Machine Learning for Data Science	Kernel Methods Mixture Models Dimensionality Reduction Sequential Models	Slides Chosen sections from Chapters 6,7,9,10,11,12, 13	Quiz 7 – October 20 Assignment 6 – October 20
9 October 23	Deep Learning for Data Science	Neural Networks Advanced Architecture	Slides Bishop: Chapter 5 Goodfellow: Chapters 6-10 chosen sections will be announced	Quiz 8 – October 27 Assignment 7 – October 27
10 October 30	Natural Language Processing & Data Science	Natural Language Processing Large Language Models	Slides Goodfellow: Chapters 6-10 chosen sections. Bird: Chapters 1-11 chosen sections will be announced	Quiz 9 – November 3 Assignment 8 – November 3
11 November 6	Computer Vision & Data Science	Convolutional Neural Networks Object Detection 3D Vision	Slides Goodfellow: Chapters 6-10 chosen sections will be announced. Szeliski: Chapters 1-5, 12, and 14 selected sections will be announced	Progress Report – November 10 Quiz 10 – November 10 Assignment 9 – November 10
12 November 13	Graph Theory & Data Science	Network Science Graphical Models Graph Representation Learning	Slides Barbasi: Chapters 1, 2, 4, 7, 9 chosen sections will be announced	Quiz 11 – November 17 Assignment 10 – November 17
13 November 20	Model Evaluation in Data Science	Decision Theory Information Theory Validation & Evaluation Experimental Design	Slides Goodfellow: Chapter 3 Bishop: Chapter 1, Sections 1.5, 1.6	Quiz 12 – November 24 Assignment 11 – November 24
14 November 27	Communication in Data Science	Project Management Development Paradigms Interpretation Ethics in Data Science	Slides	Quiz 13 – December 4 Makeup Assignment – December 8
15 December 4	Course Recap	Course Recap	Study Guide	Final Project Report Demo Code
16 December 11	Final Exam	Final Exam	Final Exam	Final Exam Part 1 Final Exam Part 2

Academic Resources

How to succeed in this class

You will receive it if you work hard, put in the effort, and use all your resources to achieve the highest grade in this class. In that regard, I recommend the following to succeed in this class:

- 1- 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.
- 2- If you are having trouble finishing an in-class assignment, that will NOT automatically result in a lousy grade. A good grade could be achieved if you showed effort and explained your thought process despite having an incorrect result.
- 3- Complete all your assignments to the best of your abilities.
- 4- See your errors and mistakes as opportunities to learn more.
- 5- Please ask questions if things need to be clarified.
- 6- Use email as the preferred way of communicating.

Technology

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/>)

Library Services

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)

Food Security

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

Mental Health

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

Writing

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/>

Tutoring

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 as accessible as possible. 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 on the basis of 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 are required to 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 you 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 which requires schools that receive federal funds to provide reasonable accommodations 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 who may need 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 percentage of similarity 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.

TWU 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. Excessive absences, even if documented, 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

Grading Policy

[https://docs.google.com/document/d/1eeTJG916awbljyMG6zIOSuak2U_ozbCN/edit?usp=drive link](https://docs.google.com/document/d/1eeTJG916awbljyMG6zIOSuak2U_ozbCN/edit?usp=drive_link)

AI Usage Policy

[https://docs.google.com/document/d/1rPm6TjS8FRFTLbWD9ERbbBySfXCcHg9j/edit?usp=drive link](https://docs.google.com/document/d/1rPm6TjS8FRFTLbWD9ERbbBySfXCcHg9j/edit?usp=drive_link)

Academic Honesty

[https://docs.google.com/document/d/1n2yJvdRV2BzUb0QezPLeZwJyUJp_eE3Z/edit?usp=drive link](https://docs.google.com/document/d/1n2yJvdRV2BzUb0QezPLeZwJyUJp_eE3Z/edit?usp=drive_link)

Instructor Policies

Holidays

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.

Diversity

The instructor is committed to diversity, inclusion, and equality in the classroom and accordance with the university policies regardless of any cultural background, country of origin, religion, race, ethnicity, and sexual orientation.

Please let the instructor know in what way 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.

Office Hours

- First come, first served.
- Maximum 15 minutes if people are waiting,
- If two or more students come at the same time, it will be in the order of the last name.
- Please come prepared to office hours with questions.

Conflict

In case of conflict between students in the classroom, the instructor will act as a mediator until proper university authorities are notified.