



CS 330 Data Science Process and Techniques

Winter 2024

Course Credits: 4

Contact Hours: 56 hours

Instructor: TBA

Email: TBA

COURSE OBJECTIVES

This course is a comprehensive course designed to equip students with the knowledge and practical skills necessary to navigate the entire data science process, from data collection and preprocessing to advanced analysis and meaningful interpretation. From formulating questions and collecting raw data to deriving actionable insights and communicating findings, this course guides students through the entire data science process.

Upon Completion of this Course, students will be able to:

1. Explore techniques for gathering data from diverse sources
2. Learn best practices for cleaning and transforming raw data into a usable format, including handling missing values and outliers
3. Utilize data visualization techniques to gain insights and identify patterns in the data
4. Explore various techniques for evaluating the performance of machine learning models and selecting the most appropriate model for a given problem
5. Work on hands-on projects that simulate real-world data science scenarios, allowing students to apply their knowledge in practical situations

PREREQUISITES

CS 212 Software Engineering; CS 296 Fundamentals of Machine Learning

GRADING

Grades will be determined by accumulating points, with 100 points being the



maximum, as follows:

ITEM	POINTS
4 Quizzes	20 Points
Midterm	30 Points
2 Projects	20 Points
Final Exam	30 Points
Total	100 Points

Late submissions will be graded at the end of the course. Grades will be assigned according to the following rule:

$$A \geq 90 > B \geq 80 > C \geq 70 > D \geq 60 > F.$$

We reserve the right to make adjustments to the overall grading policy.

COURSE MATERIALS

Required Texts:

Foster Provost; Tom Fawcett. *Data Science for Business*, 1st Edition, O'Reilly Media.

Recommended (Optional) Texts or Other Materials:

None

COURSE TOPICS

MODULE	TASKS
Module 1	<p>Topics: Topic 1: Data-Analytic Thinking Topic 2: The Ubiquity of Data Opportunities Topic 3: Data Processing and “Big Data” Topic 4: Data and Data Science Capability as a Strategic Asset</p> <p>Assessments: Quiz#1</p>



Module 2	<p>Topics: Topic 5: Business Problems and Data Science Solutions Topic 6: Data Understanding Topic 7: Other Analytics Techniques and Technologies Topic 8: Models, Induction, and Prediction</p> <p>Assessments: Quiz#2 Project#1</p>
Module 3	<p>Topics: Topic 9: Fitting a Model to Data Topic 10: Class Probability Estimation and Logistic “Regression” Topic 11: Overfitting and Its Avoidance Topic 12: Holdout Data and Fitting Graphs</p> <p>Assessments: Midterm Project#2</p>
Module 4	<p>Topics: Topic 13: Similarity, Neighbors, and Clusters Topic 14: Example: Clustering Business News Stories Topic 15: Decision Analytic Thinking I: What Is a Good Model? Topic 16: Plain Accuracy and Its Problems</p> <p>Assessments: Quiz#3</p>
Module 5	<p>Topics: Topic 17: The Confusion Matrix Topic 18: The Area Under the ROC Curve (AUC) Topic 19: Evidence and Probabilities Topic 20: Representing and Mining Text</p> <p>Assessments: Quiz#4 Final Exam</p>

ATTENDANCE

1) Class attendance is required. Missing classes without permission will lead to decrease in overall grade.

Missing less than two classes: no penalty.

Missing more than two classes: 7% will be taken off from the overall grade.

If the instructor reports a student’s frequent missing of class to the Soochow University Academic Administration Office, the student might get a written warning



and might be prohibited from attending final exam.

2) Participants in this course are expected to arrive in class promptly and adequately prepared. The primary objective of this course is to critically engage with the readings and the subject matter. Therefore, course participants are expected to have completed the reading prior to class and prepare thoughtful reflections/commentaries to share with fellow colleagues.

LEARNING REQUIREMENTS

- 1) Late assignments are not acceptable and are subjected to grade deductions.
- 2) Assignments submitted in the wrong format will be counted as not submitted.
- 3) Failure to submit or fulfill any required course component results in failure of the class.
- 4) Make-up for midterm and final exams only with valid excuses, as defined by the University.
- 5) In order to earn a Certificate of Completion, participants must thoughtfully complete all assignments by stated deadlines and earn an average quiz score of 50% or greater.

TECHNOLOGY POLICY

The use of electronic devices in class is distracting, both for the user and for the rest of the class. Only non-programmable calculators can be used in the tests and exam. Any attempts to use cell phones and other electronic communication devices will be seemed as cheating. Laptops are discouraged, unless you use them for activities DIRECTLY related to the course (eg., note taking, reading course documents).

ACADEMIC INTEGRITY POLICY

Soochow University highly values the academic integrity and aims to promote the academic fairness, honesty and responsibility. Any academic dishonesty behaviors and any attempts to cheats and plagiarism will be reported to the university administration office. A written warning and the relevant penalties will be imposed. The record might be shown on the official university transcript.



蘇州大學
Soochow University

DISABILITY ACCOMMODATION

Soochow University is committed to maintaining a barrier-free environment so that students with disabilities can fully access programs, courses, services, and activities at Soochow University. Students with disabilities who require accommodations for access to and/or participation in this course are welcome.

Note:

Please contact the University Administrative Office immediately if you have a learning disability, a medical issue, or any other type of problem that prevents professors from seeing you have learned the course material.