

CS 331: Data Science Practicum

Fall 2026

INSTRUCTOR: Dr. Brent Young

OFFICE: SLC 406

EXTENSION: 3824

E-MAIL: brent.young@wilkes.edu

WEBPAGE: <http://young.mathcs.wilkes.edu/courses/courses.html>

OFFICE HOURS: MTWF 9:00 – 9:50am, MW 1:00 – 1:50pm

Course Description

Study of practical solutions for complex problems using predictive analytics and statistical learning (supervised and unsupervised) on real-world data. Examination of the essential role of algorithms in Big Data Analysis using case studies over diverse domains. Students develop and present requirements and design plans for real-world projects.

Prerequisites

CS 231 (Introduction to Data Science) or equivalent.

Meeting Times

- Monday: 10:00 – 11:50 AM (Lab/Practicum)
- Wednesday: 10:00 – 11:15 AM (Lecture)
- Friday: 10:00 – 11:15 AM (Lecture)

Textbooks

- *Doing Data Science: Straight Talk from the Frontline* by Cathy O’Neil and Rachel Schutt
- *Data Mining and Machine Learning: Fundamental Concepts and Algorithms* by Zaki and Meira

Course Objectives

By the end of the course, students will be able to:

- Apply the data science pipeline to real-world problems,
- Perform data cleaning, exploratory data analysis, and feature engineering,
- Build and evaluate supervised and unsupervised learning models,
- Interpret and communicate results clearly in written and oral form, and
- Use version control tools to manage and document their work.

Course Modules

The course is organized into the following modules:

- **Module 1: The Data Science Workflow**

Introduction to the data science pipeline as an iterative, accountable process. Students learn to formulate a problem, define what success looks like, and set up reproducible infrastructure (version control, project organization) that will be expected throughout the semester. Introduces ethical considerations in data science.

- **Module 2: Data Wrangling and Exploratory Data Analysis**

Data cleaning, visualization, and pattern discovery on real, imperfect datasets. Emphasis is on exploration as purposeful — observations made here directly inform feature engineering in later modules.

- **Module 3: Statistical Thinking for Data Science**

Inference, uncertainty, and baseline modeling. Introduction to model evaluation strategies (cross-validation, train/validation/test splits) for honestly estimating model performance.

- **Module 4: Supervised Learning**

Classification, regression, and model evaluation. Introduction to feature engineering as a formal practice. Examine fairness and bias as they arise in predictive modeling.

- **Module 5: Text Data and Information Retrieval**

Introduction to working with unstructured text: preprocessing, representation (including feature engineering for text), and similarity methods.

- **Module 6: Unsupervised Learning and Pattern Mining**

Clustering and association analysis. Introduction to association mining as a new pattern-discovery technique.

- **Module 7: Big Data and Distributed Computing**

Introduction to the Hadoop ecosystem, HDFS, MapReduce, and PySpark as tools for working at scale. Considers how data volume changes both technical approach and ethical stakes, particularly around privacy.

- **Module 8: Model Improvement and Deployment Concepts**

Returns to the questions posed in Module 1 — now with a working model in hand. Covers advanced tuning strategies, deployment concepts, and monitoring for performance and fairness drift over time.

Assessment and Grading

- Assignments / Labs: 25%
- Semester Project: 30%
- Midterm Exam: 10%
- Final Exam: 20%
- Oral Exams (2): 10%
- Participation / Engagement: 5%

Your final grade will be determined from your raw score as follows:

Raw Score (S)	GRADE
$90\% \leq S$	A = 4.0
$85\% \leq S < 90\%$	B+ = 3.5
$80\% \leq S < 85\%$	B = 3.0
$75\% \leq S < 80\%$	C+ = 2.5
$70\% \leq S < 75\%$	C = 2.0
$65\% \leq S < 70\%$	D+ = 1.5
$60\% \leq S < 65\%$	D = 1.0
$S < 60\%$	F = 0.0

Oral Exams

Two oral exams will be conducted during the semester. These will take the form of short technical interviews in which students explain concepts, methods, and aspects of their work, with emphasis on clarity, understanding, and justification of decisions.

Project

Students will complete a semester-long project involving the full data science pipeline, culminating in a written report and presentation.

Academic Integrity

The work you hand in should be your own work. If you receive help from any source, that help must be acknowledged in writing and turned in with the assignment. If there is evidence that work you hand in is not your own, the first time you will receive a zero on the work and the second time you will receive a zero for the course grade. Appropriate deans will also be notified. See the Statement on Intellectual Responsibility and Plagiarism in the Student Handbook for more details.

Use of AI Tools

The use of AI tools (such as ChatGPT, GitHub Copilot, or similar systems) is permitted in this course as a means of support, but not as a substitute for understanding.

Any use of AI must be clearly documented. Students should indicate when and how such tools were used (e.g., for debugging, generating code suggestions, or explaining concepts).

Students are responsible for all work they submit. This includes:

- Understanding any code, analysis, or explanation included in their work
- Being able to explain and justify their choices and results
- Ensuring the correctness and appropriateness of all outputs

Work that cannot be explained by the student will be treated as a violation of academic integrity. Statements such as “an AI tool generated this” are not sufficient justification for the inclusion of material.

In this course, the goal is to develop the ability to think critically and communicate technical ideas clearly. AI tools may assist in this process, but they cannot replace it.

Attendance Policy

Attendance at all scheduled classes is expected and required. Repeated absences are a sufficient cause for failure.

After five consecutive instructional hours of unexcused absences from a class, students may be readmitted to the class only by action of the Office of Student Life and Success and the department chairperson concerned. Any absence beyond that permitted in the course is a matter between the student and the instructor. Absences due to illness, religious holidays, or participation in athletic or other University sponsored activities are usually considered to be acceptable reasons for absences, but notification of such absences and arrangements to make up missed work should be made with the instructor by the student.

In the unfortunate event of a death in the family, students are asked to contact the Office of Student Life and Success so that notification might be sent to faculty members and arrangements made with them to assist students in making up missed work.

If students are ill and will be missing a test, examination, or presentation, it is their responsibility to contact the instructor by email prior to the start of the missed activity. When students are going to be absent for a period of two days or more, if they notify the Office of Student Life and Success, written notification of their extended absence will be sent to the students' instructors.

Notice of Nondiscrimination

Wilkes University is committed to providing a welcoming environment for all members of our community and to ensuring that all educational and employment decisions are based on individuals' abilities and qualifications.

Wilkes University prohibits discrimination in its educational programs, employment, admissions or any activities on the basis of race, color, national or ethnic origin, age, religion, disability, pregnancy, gender, gender identity and/or expression, sexual orientation, marital or family status, military or veteran status, genetic information, or any other characteristic protected under applicable federal, state or local laws. Discriminatory conduct including sexual harassment and other sexual misconduct or violence such as rape, sexual assault, sexual exploitation and coercion will not be tolerated.

Consistent with this principle, Wilkes University will comply with state and federal laws such as the Pennsylvania Human Relations Act or other applicable state law, Title IX, Title VI and Title VII of the Civil Rights Act, the Americans with Disabilities Act, Section 504 of the Rehabilitation Act of 1973, the Age Discrimination in Employment Act of 1967, the Ethnic Intimidation Act of 1982 (P.L. 537-154) and other laws that prohibit discrimination.

Any member of the Wilkes University community has the right to raise concerns or make a complaint regarding discrimination under this policy without fear of retaliation. Inquiries about this policy statement may be addressed to the University's Title IX coordinator. Anonymous complaints may also be made by filing a report online.