Northeastern
College of Engineering

Info 6105 Data Sci Eng Methods

FALL 2025

Institution: Northeastern University

Semester: Fall 2025

Instructor Details

- **Name:** Professor William Claster
- **Email:** w.claster@northeastern.edu
- **Office Hours:** Wednesday 12:30 pm to 2:30 pm
- **Contact:** 626-376-5444

Instructor Biography

William Claster is an Associate Professor of Data Science with deep expertise in the field, having authored two books on data science. He has a diverse research portfolio, including work in natural language processing (NLP), the intersection of NLP and medicine, and social media analytics. Dr. Claster has lived in Japan for many years, which has enriched his global perspective and approach to data science. His work bridges the gap between theoretical research and practical applications, making significant contributions to both academia and industry.

Course Prerequisites

- Knowledge of Python

Course Overview

The Data Science and Engineering Tools course introduces students to both foundational machine learning techniques and cutting-edge **Agentic AI systems**. Students will learn to build

autonomous AI agents using LangChain and LangGraph, develop **Multi-Agent Systems (MAS)**, and implement **RAG (Retrieval-Augmented Generation)** architectures alongside traditional data science methods. The course emphasizes practical implementation of both agentic systems and classical ML algorithms, preparing students for the evolving landscape where AI agents augment traditional data pipelines.

Core topics include:

- **Agentic AI**: LangChain fundamentals, LangGraph state machines, Multi-Agent Systems, RAG architectures
- **Supervised Learning**: parametric and nonparametric algorithms, support vector machines, neural networks
- **Unsupervised Methods**: clustering, dimensionality reduction, recommender systems
- **Deployment**: Docker containerization and Streamlit applications

By the end of the course, students will be equipped to build sophisticated Multi-Agent Systems, analyze datasets, design predictive models, and deploy end-to-end solutions for real-world challenges.

Learning Objectives

- Design and implement autonomous AI agents using LangChain/LangGraph
- Build RAG systems for knowledge-augmented AI applications
- Orchestrate multi-agent systems for complex problem solving
- Understand foundational and advanced concepts of machine learning models
- Develop practical skills in data exploration, visualization, and modeling
- Gain experience working on real-world datasets to solve data science problems
- Understand when to use agentic vs. traditional ML approaches

---

Course Outline

Tentative Week-by-Week Breakdown

Week 1 Introduction & Agentic AI Fundamentals • Introduction to Data Science field • LangChain basics and agent types • Python setup and foundational libraries • Math for Data Science I

Week 2 LangGraph & Data Exploration • Building stateful agents with LangGraph • Data exploration with Pandas, Matplotlib, Seaborn • Math for Data Science II

Week 3 RAG/Multi-Agent Systems & KNN • RAG architecture introduction • Multi-agent coordination basics • KNN, Linear regression • Training/Testing/Validation splits **Assignment 1 Due**

Week 4 Modeling & Classification • Decision trees, evaluation metrics • Basic classification algorithms • KL Divergence and Cross Validation

Week 5 Logistic Regression • Logistic regression theory and implementation • Model evaluation and interpretation

Week 6 Unsupervised Learning • Clustering methods • Resampling (Cross-Validation, Bootstrap) • Ensembles and Super learner **Assignment 2 Due**

Week 7 Bias-Variance & Generalization • Bias-variance tradeoff • Bayes and Markov Chains • Model generalization

Week 8 MIDTERM EXAM • Math for Data Science III • Time Series, Moving Averages • Review session first half of class

Week 9 Attribute Engineering • Feature engineering techniques • Hyperparameter optimization • Feature selection methods **Assignment 3 Due**

Week 10 Regularization • L1/L2 regularization • Overfitting prevention techniques

Week 11 Ridge and Lasso Regression • Ridge regression • Lasso regression • Elastic Net

Week 12 Support Vector Machines • SVM theory • Kernel methods • SVM implementation **Assignment 4 Due**

Week 13 PCA & Final Projects • Principal Component Analysis • Dimensionality reduction • Final project presentations **Final Project Due**

Week 14 Neural Networks & Review • Neural network fundamentals • Deep learning introduction • Comprehensive review for final exam

Week 15 FINAL EXAM WEEK • Comprehensive final examination

Grading Policy

Component	Weight	Details
Assignments	30%	4 assignments × 7.5% each

Component	Weight	Details
Midterm Exam	15%	Week 8 (1 hour, in-person, paper, pen, 1 formula sheet)
Final Exam	20%	Week 15 (2 hours, in-person, comprehensive, paper, pen, 1 formula sheet)
Final Project	15%	Due Week 13
Video Assignments on Edpuzzle	10%	Various topics
Class Participation	10%	Q&A, discussions, peer engagement

Assignment Details

- **Assignment 1 (Week 3):** Build a basic LangChain agent with tools
- **Assignment 2 (Week 6):** Implement ML algorithm from scratch + sklearn comparison OR "Compare different sklearn models on a real dataset" OR Create a data preprocessing pipeline with pandas
- **Assignment 3 (Week 9):** Create a RAG system OR ensemble classifier
- **Assignment 4 (Week 12):** Integrate agents with traditional ML pipeline

Examination Topics

Midterm Exam (Week 8) - Covers Weeks 1-7:

- Agentic AI: LangChain components, LangGraph workflows, agent types, RAG fundamentals
- Traditional ML: KNN, regression, decision trees, cross-validation, bias-variance
- Integration: When to use agents vs. traditional ML

Final Exam (Week 15) - Comprehensive, emphasis on Weeks 8-14:

- Advanced ML: Regularization, SVM, PCA, neural networks, ensembles
- Advanced Agentic AI: Complex RAG, memory systems, production deployment
- System design and real-world applications

Resources

Required Tools/Software

- Python (latest version)
- **AI/Agent Libraries:** LangChain, LangGraph, OpenAI/Anthropic API
- **ML Libraries:** Scikit-learn, Pandas, NumPy, Matplotlib
- **Vector Databases:** Pinecone/Chroma/FAISS
- Jupyter Notebooks and Python scripts

- Docker for containerization
- Streamlit for GUI development
- GitHub for version control

Recommended Reading/References

- *Python for Data Analysis* by Wes McKinney
- *Hands-On Machine Learning* by Aurélien Géron
- *Introduction to Statistical Learning* (https://www.statlearning.com)
- LangChain Documentation (https://python.langchain.com)
- LangGraph Documentation

Policies

Attendance

Weekly attendance is strongly encouraged. Active participation is critical for success.

Submission Deadlines

Assignments must be submitted on the due date. Late submissions penalized 10% per day.

Plagiarism

Zero tolerance for copied work. Projects must reflect individual effort. Academic integrity strictly enforced.

Accessibility

Support for students with disabilities provided per university policies.

Key Dates

Event	Week
Assignment 1 Due	End of Week 3
Assignment 2 Due	End of Week 6
Midterm Exam	Week 8 (in-class)
Assignment 3 Due	End of Week 9
Assignment 4 Due	End of Week 12

Event	Week
Final Project Due	Week 13
Study Week	Week 14
Final Exam	Week 15 (finals week)

Grading Scale

Grade	Percentage	Grade	Percentage	Grade	Percentage
A	95-100%	B+	87-89.9%	C+	77-79.9%
A-	90-94.9%	B	84-86.9%	C	74-76.9%
		B-	80-83.9%	C-	70-73.9%
				F	Below 70%

University Policies

Attendance/Late Work Policy

Students registered in MGEN courses (INFO, CSYE, and DAMG) are allowed a maximum of 2 absences per course, with 3 or more absences resulting in an automatic 'F' for that course. Students must inform instructors of absences in advance. For medical issues preventing attendance, contact Academic Advisor (coe-mgen-gradadvising@northeastern.edu).

Teaching Assistants (TAs) or Instructional Assistants (IAs) will collect attendance at each class.

Students must submit assignments by the deadline. Communication with faculty required prior to deadline for late submissions. Work submitted late without prior communication will not be graded.

End-of-Course Evaluation Surveys

Your feedback is important to the College of Engineering. Complete the evaluation survey at https://neu.evaluationkit.com. Responses are completely anonymous and confidential.

Academic Integrity

A commitment to academic integrity is essential to Northeastern University's mission. Students must know requirements regarding research, writing, examinations, collaborative work, and use of study aids. Students must complete all work according to the highest ethical standards.

Full policy: http://www.northeastern.edu/osccr/academic-integrity-policy/

MGEN Student Feedback

Provide anonymous feedback using:
https://neu.co1.qualtrics.com/jfe/form/SV_cTIAbH7ZRaaw0Ki

University Health and Counseling Services

For health issues, contact UHCS: https://www.northeastern.edu/uhcs

Student Accommodations

The Disability Resource Center (DRC) provides services for qualifying students under Section 504 and ADAAA. Visit: https://drc.sites.northeastern.edu

Library Services

Resources include over 900,000 print volumes, 206,500 e-books, and 70,225 electronic journals. Visit: https://library.northeastern.edu

24/7 Canvas Technical Help

Call: 617-373-4357 or email: help@northeastern.edu Canvas Resources:
https://canvas.northeastern.edu/student-resources/

Diversity and Inclusion

Northeastern is committed to equal opportunity, affirmative action, diversity, and social justice. Visit: http://www.northeastern.edu/oidi/

Title IX

Title IX protects from sex or gender-based discrimination in educational programs. Emergency: Call 911 Resources: https://www.northeastern.edu/ouec