Northeastern
College of Engineering

INFO 7390 Advances in Data Science and Architecture

Course Information
Course Title: Advances in Data Science and Architecture
Course Number: INFO 7390
Term and Year: Fall 2025
Credit Hour: 4
CRN: 17739
Course Format: On-Ground

Instructor Information
Full Name: Junwei Huang
Email Address: jun.huang@northeastern.edu
Office Hours: TBA

Instructor Biography
Dr Junwei Huang is a seasoned data and machine learning scientist with 15+ years of industry experience and a proven record of instructing graduate level data science and machine learning courses, deploying machine learning models, and delivering scalable data ETL pipelines. Dr Huang earned a PhD in Geophysics and a MSc in Computer Science.

Teaching Assistant Information
Full Name: TBA
Email Address: TBA
Office Hours: TBA

Course Prerequisites
- Graduate level **INFO 6105** Minimum Grade of B-
- Undergraduate level INFO 5100 Minimum Grade of B- or Graduate level INFO 5100 Minimum Grade of B- or Graduate level CSYE 6200 Minimum Grade of B-

Course Description
This proposed course is designed to equip students with the practical skills and knowledge required to excel in the field of data science and machine learning. The topics include data acquisition, web scraping, data cleaning, machine learning model training, testing, deployment, and monitoring. Additionally, it will delve into various areas such as supervised learning, unsupervised learning, recommendation systems, anomaly detection, natural language processing and other relevant topics in the industry.

After taking this course, students will be equipped with theoretical understanding and hands-on projects and understand how to present their experiences on resume and in interviews. The course includes individual assignments and group project.

Textbook (Optional)

- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems by Geron, Aurelien
- Web Scraping with Python: Collecting More Data from the Modern Web by Mitchell, Ryan
- Speech and Language Processing by Jurafsky
- Deep Learning by Goodfellow, Ian

Standard Learning Outcomes

Upon successful completion of the course, students will:

- Acquire a comprehensive understanding of the data science lifecycle, from data acquisition to model deployment and monitoring.
- Demonstrate proficiency in data acquisition techniques and web scraping to collect data from various sources.
- Apply effective data cleaning and preprocessing methods to handle missing data, outliers, and feature engineering for model building.
- Develop a strong foundation in supervised learning algorithms, including regression and classification, and effectively evaluate and select models.
- Gain proficiency in unsupervised learning techniques such as clustering and dimensionality reduction for pattern discovery and data exploration.
- Understand the concepts and techniques used in recommendation systems, including collaborative filtering and content-based filtering.
- Apply anomaly detection methods to identify and handle abnormal patterns in data for various domains and applications.
- Demonstrate the ability to train, test, and deploy machine learning models while considering model performance, evaluation metrics, and best practices.
- Develop skills in monitoring and maintaining deployed models, including detecting concept drift, model retraining, and addressing ethical considerations.
- Understand job interviews process, and gain insights into industry trends and job market expectations.

Grading/Evaluation Standards

Grade Scale

95-100%	**A**	**87-89.9%**	**B+**	**77-79.9%**	**C+**	**69.9% or below**	**F**
		84-86.9%	**B**	**74-76.9%**	**C**		
90-94.9%	**A-**	**80-83.9%**	**B-**	**70-73.9%**	**C-**		

Grade Breakdown:

This course consists of 4 projects + 1 deep dive.

- 15%: Web Data Extraction / Web scraper
- 15%: Deployment of Supervised learning Model
- 20%: Recommender system
- 20%: Anomaly detection
- 20%: One NLP topic: Sentiment Analysis, Named Entity Recognition
- 10%: Extension of a previous project + Oral presentation + Social media style write-up

Course Schedule

The course schedule similar to the one below will be followed in class.

1	Introduction to Applied Data Science	• Overview of data science and its applications • Introduction to key concepts, tools, and techniques
2	Data Acquisition and Web Scraping	• Understanding data sources and formats • Techniques for web scraping and data collection
3	Data Cleaning and Preprocessing	• Exploratory data analysis (EDA) and data visualization techniques • Handling missing data, outliers, and data imputation methods
4	Supervised Learning: Regression and Classification	• Introduction to regression and classification algorithms • Model selection, evaluation, and hyperparameter tunin
5	Unsupervised Learning: Clustering and Dimensionality Reduction	• Overview of clustering algorithms • Dimensionality reduction techniques and evaluation metrics
6	Model Training, Testing, and Deployment	• Model training techniques and best practices • Model evaluation metrics and cross-validation
7	Model Monitoring and Maintenance	• Techniques for monitoring model performance and concept drift detection • Strategies for model retraining and updating
8	Natural Language Processing (NLP) Fundamentals	• Introduction to NLP and its applications • Text preprocessing techniques and tokenization
9	NLP: Text Classification, Sentiment Analysis, and Named Entity Recognition	• Techniques for text classification and sentiment analysis • Named Entity Recognition (NER) techniques
10	NLP: Text Generation and Language Models	• Text generation using language models • Introduction to language models like GPT-3
11	Recommendation Systems	• Collaborative filtering techniques • Content-based filtering methods
12	Anomaly Detection	• Introduction to anomaly detection techniques • Statistical approaches and machine learning-based methods
13	Ethical Considerations and Bias Detection	• Ethical considerations in data science and machine learning • Bias detection and mitigation in deployed models
14	Interview Preparation and Career Guidance	• Mock interviews demonstration • Effective communication of technical concepts • Industry trends, job market insights, and networking opportunities
15	Final Project Presentations.	Oral presentation

Attendance/Late Work Policy

Attendance 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 are expected to inform their instructors of any absences in advance of the class; if a student is sick long-term or experiences a medical issue that prevents class attendance, it is strongly encouraged that they speak with their Academic Advisor (coe-mgen-gradadvising@northeastern.edu) to learn more about the Medical Leave of Absence. Should a student anticipate being unable to attend 3 or more classes, they should discuss their situation with their Academic Advisor to explore other types of leave in accordance with the University's academic and global entry expectations. International students should review the Office of Global Services webpage to understand their visa compliance requirements.

Teaching Assistants (TAs) or Instructional Assistants (IAs) will be present at each class to collect student attendance.

Late Work Policy

Students must submit assignments by the deadline in the time zone noted in the syllabus. Students must communicate with the faculty prior to the deadline if they anticipate work will be submitted late. Work submitted late without prior communication with faculty will not be graded.

End-of-Course Evaluation Surveys
Your feedback regarding your educational experience in this class is particularly important to the College of Engineering. Your comments will make a difference in the future planning and presentation of our curriculum.

At the end of this course, please take the time to complete the evaluation survey at **https://neu.evaluationkit.com**. Your survey responses are **completely anonymous and confidential**. For courses 6 weeks in length or shorter, surveys will be open one week prior to the end of the courses; for courses greater than 6 weeks in length, surveys will be open for two weeks. An email will be sent to your Northeastern University Mail account notifying you when surveys are available.

Academic Integrity
A commitment to the principles of academic integrity is essential to the mission of Northeastern University. The promotion of independent and original scholarship ensures that students derive the most from their educational experience and their pursuit of knowledge. Academic dishonesty violates the most fundamental values of an intellectual community and undermines the achievements of the entire University.

As members of the academic community, students must become familiar with their rights and responsibilities. In each course, they are responsible for knowing the requirements and restrictions regarding research and writing, examinations of whatever kind, collaborative work, the use of study aids, the appropriateness of assistance, and other issues. Students are responsible for learning the conventions of documentation and acknowledgment of sources in their fields. Northeastern University expects students to complete all examinations, tests, papers, creative projects, and assignments of any kind according to the highest ethical standards, as set forth either explicitly or implicitly in this Code or by the direction of instructors.

Go to http://www.northeastern.edu/osccr/academic-integrity-policy/ to access the full academic integrity policy.

MGEN Student Feedback
Students who would like to provide the MGEN unit with anonymous feedback on this particular course, Teaching Assistants, Instructional Assistants, professors, or to provide general feedback regarding their program, may do so using this survey: https://neu.co1.qualtrics.com/jfe/form/SV_cTIAbH7ZRaaw0Ki

University Health and Counseling Services
As a student enrolled in this course, you are fully responsible for assignments, work, and course materials as outlined in this syllabus and in the classroom. Over the course of the semester if you experience any health issues, please contact UHCS.

For more information, visit https://www.northeastern.edu/uhcs.

Student Accommodations
Northeastern University and the Disability Resource Center (DRC) are committed to providing disability services that enable students who qualify under Section 504 of the Rehabilitation Act and the Americans with Disabilities Act Amendments Act (ADAAA) to participate fully in the activities of the university. To receive accommodations through the DRC, students must provide appropriate documentation that demonstrates a current substantially limiting disability.

For more information, visit https://drc.sites.northeastern.edu.

Library Services
The Northeastern University Library is at the hub of campus intellectual life. Resources include over 900,000 print volumes, 206,500 e-books, and 70,225 electronic journals.

For more information and for education specific resources, visit https://library.northeastern.edu
Network Campus Library Services: Northeastern University Library Global Campus Portals

24/7 Canvas Technical Help
For immediate technical support for Canvas, call 617-373-4357 or email help@northeastern.edu

Canvas Student Resources: https://canvas.northeastern.edu/student-resources/

For assistance with my Northeastern e-mail, and basic technical support:
Visit ITS at https://its.northeastern.edu
Email: help@northeastern.edu
ITS Customer Service Desk: 617-373-4357

Diversity and Inclusion
Northeastern University is committed to equal opportunity, affirmative action, diversity, and social justice while building a climate of inclusion on and beyond campus. In the classroom, members of the University community work to cultivate an inclusive environment that denounces discrimination through innovation, collaboration, and an awareness of global perspectives on social justice.

Please visit http://www.northeastern.edu/oidi/ for complete information on Diversity and Inclusion

Title IX
Title IX of the Education Amendments of 1972 protects individuals from sex or gender-based discrimination, including discrimination based on gender-identity, in educational programs and activities that receive federal financial assistance.

Northeastern's Title IX Policy prohibits Prohibited Offenses, which are defined as sexual harassment, sexual assault, relationship or domestic violence, and stalking. The Title IX Policy applies to the entire community, including male, female, transgender students, faculty, and staff.

In case of an emergency, please call 911.

*Please visit **https://www.northeastern.edu/ouec** for a complete list of reporting options and resources both on- and off-campus.*