



Northeastern University

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 2024

Credit Hour: 4

Course Format: On-Ground

Instructor Information

Full Name: Junwei Huang

Email Address: jun.huang@northeastern.edu

Course Prerequisites

Graduate level **INFO 6105** Minimum Grade of B- and 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

Learning outcomes common to all College of Engineering Graduate programs:

1. *An ability to identify, formulate, and solve complex engineering problems.*
2. *An ability to explain and apply engineering design principles, as appropriate to the program's educational objectives.*
3. *An ability to produce solutions that meet specified end-user needs with consideration of public health, safety, and welfare, as well as global, cultural, social, environmental, and economic factors.*



Northeastern University

College of Engineering

The Information Systems Program accepts students of different engineering backgrounds with minimum programming skills and produces first class Information Systems engineers that operate at the intersection of real-world complexity, software development, and IT management. Graduating students will be able to construct end-to-end advanced software applications that meet business needs.

Specific Learning Outcomes for the Information Systems program:

1. Create a strong technical foundation through diverse, high-level courses
2. Built crucial interpersonal skills needed to succeed in any industry
3. Foster a deep level of applied learning through project-based case studies

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



Northeastern University

College of Engineering

- 2%: (Bonus) In-class attendance and Canvas discussion

Course Schedule

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

1	Introduction to Applied Data Science	<ul style="list-style-type: none"> • Overview of data science and its applications • Introduction to key concepts, tools, and techniques
2	Data Acquisition and Web Scraping	<ul style="list-style-type: none"> • Understanding data sources and formats • Techniques for web scraping and data collection
3	Data Cleaning and Preprocessing	<ul style="list-style-type: none"> • Exploratory data analysis (EDA) and data visualization techniques • Handling missing data, outliers, and data imputation methods
4	Supervised Learning: Regression and Classification	<ul style="list-style-type: none"> • Introduction to regression and classification algorithms • Model selection, evaluation, and hyperparameter tuning
5	Unsupervised Learning: Clustering and Dimensionality Reduction	<ul style="list-style-type: none"> • Overview of clustering algorithms • Dimensionality reduction techniques and evaluation metrics
6	Model Training, Testing, and Deployment	<ul style="list-style-type: none"> • Model training techniques and best practices • Model evaluation metrics and cross-validation
7	Model Monitoring and Maintenance	<ul style="list-style-type: none"> • Techniques for monitoring model performance and concept drift detection • Strategies for model retraining and updating
8	Natural Language Processing (NLP) Fundamentals	<ul style="list-style-type: none"> • Introduction to NLP and its applications • Text preprocessing techniques and tokenization
9	NLP: Text Classification, Sentiment Analysis, and Named Entity Recognition	<ul style="list-style-type: none"> • Techniques for text classification and sentiment analysis • Named Entity Recognition (NER) techniques
10	NLP: Text Generation and Language Models	<ul style="list-style-type: none"> • Text generation using language models • Introduction to language models like GPT-3
11	Recommendation Systems	<ul style="list-style-type: none"> • Collaborative filtering techniques • Content-based filtering methods
12	Anomaly Detection	<ul style="list-style-type: none"> • Introduction to anomaly detection techniques • Statistical approaches and machine learning-based methods
13	Ethical Considerations and Bias Detection	<ul style="list-style-type: none"> • Ethical considerations in data science and machine learning • Bias detection and mitigation in deployed models
14	Interview Preparation and Career Guidance	<ul style="list-style-type: none"> • Mock interviews demonstration • Effective communication of technical concepts • Industry trends, job market insights, and networking opportunities
15	Final Project Presentations.	Oral presentation

Teaching Courseware

The teaching courseware is that the Professor teach in the class, includes the slides, code files and data sets if needs.

The courseware will be posted after the class mostly within the same day, no later than the next day. You can find the courseware information at Canvas => Modules.

End-of-Course Evaluation Surveys

Your feedback regarding your educational experience in this class is very important to the College of Professional Studies. 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,



Northeastern University

College of Engineering

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 Husky 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.

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 <http://www.northeastern.edu/drc/getting-started-with-the-drc/>.

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 <http://subjectguides.lib.neu.edu/edresearch>.

24/7 Canvas Technical Help

For immediate technical support for **Canvas**, call 1 (833) 450-3937.



Within Canvas, click **Help** on the left panel and open up links for your specific cases including: [Canvas Readiness Checklist](#), [Chat with 24/7 Canvas Support \(Faculty\)](#), [Chat with 24/7 Canvas Support \(Students\)](#), [Call 24/7 Canvas Support](#): 1 (833) 450-3937, [Access Northeastern Tech Support](#), and many more options.

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, member of the University community



Northeastern University

College of Engineering

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 www.northeastern.edu/titleix for a complete list of reporting options and resources both on- and off-campus.