INFO 6105: Data Science Engineering Methods and Tools
Fall 2025

Course Information
Course Title: Data Science Engineering Methods and Tools
Course Number: INFO6105
Term and Year: Fall 2025
Credit Hour: 4.0
CRN: 17641
Course Format: Traditional

Instructor Information
Full Name: Yizhen Zhao
Email Address: yizhenzhao@northeastern.edu
Office Hours: One Hour After Each Lecture Weekly

Instructor Biography
Dr. Yizhen Zhao received her PhD from Johns Hopkins University. Before she joined Northeastern University, she had been working as a quant specialized in machine learning, algorithm trading and financial risk modeling for years. Dr. Yizhen Zhao's research and work experience cover both the buy-side and the sell-side finance, which allows her to incorporate Wall Street practice into the classroom for students. Her teaching philosophy emphasizes project-based learning. In her classes, students work on real-world problems to advance their knowledge in data science, fin-tech and software engineering.

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

Course Prerequisites
There is no prerequisite for the course, but students are expected to be comfortable with statistics, matrix analysis and at least one type of programming language. Much of the course addresses machine learning algorithms, and students who are deficient in statistics may find it difficult to keep up. I will do my best to teach the course at a level appropriate for someone who has not had any instruction in finance. Some background material may be taught using differential equations and stochastic calculus; this material will only be used for the demonstration of concepts, and you will not be tested on it. I will make myself available before and after classes if you would like to meet to further discuss class material or related topics.

Course Description
Machine learning has had fruitful applications on Wall Street well before the advent of mobile banking apps, proficient chatbots, or search engines. This class is designed to teach introductory machine learning

techniques as applied in trading and finance. The course combines methodology with theoretical foundations and computational aspects. The focus of the course is on implementation rather than theory. It treats both the art of financial intuition and the science of realizing algorithms in Python. Students will learn how to implement algorithms and to solve real-world problems faced by quants, market makers and portfolio managers. The course will guide students through five case studies and applications like running a competition. Topics include: (i)supervised learning (parametric/non-parametric methods, GARCH, Kalman Filter, linear regressions and model selection methods, basic neural networks, decision trees, random forests, xgboost, etc.). (ii) unsupervised learning (kernel density estimation, mixture modeling, clustering, isolation forest, etc.). (iii) best practices in machine learning (pattern recognition, natural language processing basics, CNN, etc.).

Course Learning Outcomes

1) This class session focuses on the domain of machine learning in finance. The session will cover major machine learning algorithms in the context of finance using Python. The course will also introduce leading quantitative models used by finance professionals. The focus of the course will be on implementation rather than on theories. After learning this course, students are expected to master the mainstream machine learning algorithms; understand the implications of algorithm building blocks spanning across matrix analysis, statistics, optimization, econometrics and stochastic calculus; and obtain hands-on experience in working with quantitative finance modeling using Python.
2) Proficiency in Python Data Science Fundamentals: Students will attain proficiency in fundamental concepts of data science using Python, including numerical analysis, scientific computing, and manipulation of large datasets.
3) Data Analysis and Manipulation: Students will develop advanced skills in data manipulation and analysis using Python libraries to be able to support large and high-dimensional datasets, performing group and pivot operations, and managing missing data efficiently.
4) Effective Data Visualization Techniques: Students will master data visualization techniques using Python libraries to create visually appealing and informative plots to communicate insights effectively from data analysis.

Required Tools and Course Textbooks.

Python

Course Schedule/Topics Covered. The schedule presented here is tentative and subject to change. Some topics may take more or less time than anticipated. Additional topics may be included as time permits and/or according to the students' preference.

Week	In Class Topic	Assignment Due
1	lecturer introduction + students self-introduction + roadmap of the class	TBD on an ongoing basis
2	statistical distributions: from uniform to multivariate normal + grouping draw-lot	TBD on an ongoing basis
3	bootstrap, monte-carlo + mixture density + in class coding competition	TBD on an ongoing basis
4	basics about Market Making, parameter estimation, statistical tests + 1st in class	TBD on an ongoing basis

	presentation	
5	GARCH + Kalman Filter	TBD on an ongoing basis
6	coding competition – parameter fine tuning in gradient descent method newton method + quasi newton method + BFGS	TBD on an ongoing basis
7	statistical arbitrage: cointegration + pairs trading linear regression: anatomy	TBD on an ongoing basis
8	SPRING BREAK – NO CLASSES	
9	linear regression: pitfalls + supervised machine learning: Ridge Regression, LASSO, Elastic Net, Least Angle Regression (LARS) feature engineering 1: momentum factors + candlesticks	TBD on an ongoing basis
10	feature engineering 2: value factors, CAPM + Fama French 3 Factor + basic neural network model with trading signals in-class coding competition	TBD on an ongoing basis
11	Decision-tree methods: basics, feature importance, RFECV, random forest, Ada boosting + gradient boosting	TBD on an ongoing basis
12	in-class coding competition and presentation: visualize feature importance, feature selection with RFECV, run your first random forest and xgboost classifier / regression algorithm.	TBD on an ongoing basis
13	Unsupervised machine learning clustering + isolation forest	TBD on an ongoing basis
14	NLP and Anti-Money Laundering Natural Language Processing: vectorization of text and images Convolutional Neural Networks (CNN)	TBD on an ongoing basis
15	Final Exam	TBD on an ongoing basis

Assignment Grading

Three Quizzes	15%
Final Exam	35%
Projects and Assignments Presentation	30%
Coding Competition (Class Participation)	10%
Attendance	10%
Extra Credit	5%

Grading Scale

	87-89.9% B+	77-79.9% C+	

95-100% A	84-86.9% B	74-76.9% C	
90-94.9% A-	80-83.9%B-	70-73.9% C-	69.9% or below F

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_cTlAbH7ZRaaw0Ki

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.*