INFO 6105 INFO Data Science Engineering Methods and Tools
[FALL 2024]

Course Information
Course Title: Data Science Engineering Methods and Tools
Course Number: INFO 6105
Term and Year: Fall 2024
Credit Hour: 4
CRN: 21534
Course Format: On-Ground
Timing: Thursday's 1:30 – 4:50 pm PST.
Lecture location: 1425 (Vancouver Campus)

Instructor Information
Full Name: Rushdi Alsaleh, PhD
Email Address: r.alsaleh@northeastern.edu
Office Hours: On-campus:
Wed 12-2 pm
Thursday 12-1:30 pm
Friday 4:30-5:30 pm
or
Online through the link:
https://calendly.com/ralsaleh/15-min-gn
on Monday's and Tuesday's
Or
Anytime by appointment

Instructor Biography
Professor of Data Science, and Artificial Intelligence, with more than 12 years of experience in industry and has led multiple high-impact data-driven projects with local and global companies.

Teaching Assistant Information
NA

Course Prerequisites
Please consult school listings for prerequisites.

Course Description

This course provides an introduction to machine learning using Python, the open source, programming language extensively adopted by the machine learning community and industry. Machine learning is at the center of a powerful movement. Many industries depend on practitioners of machine learning to create products that parse, reduce, simplify and categorize data, and then extract actionable intelligence from that data. Professionals who are familiar with machine learning, a key technology driving Big Data, secure a competitive edge in exciting careers in the data sciences. In this course, you will use Python to learn machine learning concepts, terms and methodology, and gain an intuitive understanding of the mathematics underlying it by building actual applications. The techniques you'll learn can be a starting point to build real-world applications such as search engines, image analysis, bioinformatics, industrial automation, speech recognition, and more.

This course establishes a basic understanding of supervised learning and Bayesian classifiers using the histogram as a starting point. It then covers the design and application of practically useful classifiers such as k-nearest neighbors, linear machines and decision trees. You will also learn concepts in unsupervised learning and clustering algorithms such as expectation maximization and k-means clustering. The course concludes with the application of neural networks in machine learning.

The course uses examples to guide you through foundational concepts, often employing live algorithms to facilitate visual understanding. Pseudocode will be provided for most of the algorithms covered. You are encouraged to use the pseudocode as a reference to create your own programs in Python. The class has in-class quizzes to gauge learning and group activities including discussion. Homework assignments involving programming in Python are designed for in-depth practice.

Course 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.

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

Required Tools and Course Textbooks.
The material for this course is designed from sctarch and the following textbooks are recommended:

Recommended Textbooks:

- *Data Mining for Business Analytics: Concepts, Techniques and Applications in Python*. 2019. Galit Shmueli, Peter C. Bruce, Peter Gedeck, Nitin R. Patel. ISBN: 978-1-119-54984-0
- *An introduction to statistical learning: With applications in python*. James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J., Springer Nature.
- *Data Mining: Concepts and Techniques*, 3rd ed. The Morgan Kaufmann Series in Data Management Systems Morgan Kaufmann Publishers, 2011
- *The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second Edition*. Hastie, T., Tibshirani, R., & Friedman, J. 2009, Springer Nature.
- *Machine Learning*, Tom M. Mitchell, McGraw-Hill, 1997, ISBN-10: 0070428077, ISBN-13: 978-0070428072.
- *Pattern Recognition and Machine Learning*, Christopher M. Bishop, Springer, 2007, ISBN-10: 0387310738, ISBN-13: 978- 0387310732.
- *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd Edition, Ian H. Witten, et al., Morgan Kaufmann, 2011, ISBN-10: 0123748569, ISBN-13: 978-0123748560.
- *Data Science Using Python and R*. Chantal D. Larose, Daniel T. Larose, 2019. ISBN: 978-1-119-52681-0. Wiley

Software

- **Python Anaconda**
  https://www.continuum.io/anaconda-overview
- **R (Statisical programming language)**
  https://www.r-project.org/
- **RStudio (IDE)**
  https://posit.co/download/rstudio-desktop/

Course Schedule/Topics Covered.

Week	Date	In Class Topic	Assignment Due
1	**09/05**	**Introduction: Concepts in Machine Learning and Predictive Analytics**	**Python certificate for Data Science (DS) (Module 1)**
2	**09/12**	**Intro to python/R and Data Pre-processing, Data Visualization, EDA**	**Python Certificate for DS (Module 2)**
3	**09/19**	**Linear Models, Model Inference and Interpretation, Regression, Logistic Regression**	**Python Certificate for DS (Module 3)** **Assignment 1**
4	**09/26**	**Supervised Learning: Decision Tree, CART, Bagging and Random Forest**	**Python Certificate for DS (Module 4)**

5	**10/3**	**Supervised Learning: Decision Tree, CART, Bagging and Random Forest** **Supervised Learning: Support Vector Machine (SVM)**	**Python Certificate for DS (Module 5)** **Assignment 2**
6	**10/10**	**Supervised Learning: Support Vector Machine (SVM)** **Case studies**	**Assignment 3** **Python Certificate for DS Submission**
7	**10/17**	**Midterm Exam**	**Midterm Exam** **Python Certificate for Data Visualization (DV) (module 1)**
8	**10/24**	**Research progress report** **Neural Netwrks theories and implementation**	**Python Certificate for DV (module 2)**
9	**10/31**	**Neural Netwrks theories and implementation** **Unsupervised learning: finding Groups of Data (Clustering)**	**Assignment 4** **Python Certificate for DV (module 3)**
10	**11/7**	**Unsupervised learning: finding Groups of Data (Clustering)**	**Python Certificate for DV (module 4)**
11	**11/14**	**Evaluating and Improving Model Performance** **Overfitting and Regularization** **Applying Machine Learning Guidance and Practical Issues**	**Assignment 5** **Python Certificate for DV (module 5)**
12	**11/21**	**Classification: Nearest Neighbors, Naïve Bayes** **Applying Machine Learning Guidance and Practical Issues**	**Python Certificate for DV Submission**
13	**11/28**	**Into to deep learning**	
14	**12/5**	**Final Term Project Presentation**	**Project Presentation**
15	**12/12**	**Final Term Project Presentation**	**Project Presentation**
		Final Grade submission Deadline (Dec. 16th)	

Assignment Grading

Please insert all assignment grades and weights for the course. Example below:

- Attendance and participation – 5%
- Pyhton Certificates – 10%
- Assignments – 30%
- Midterm Exam – 25%
- Term Project – 30%

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) or the Instructor 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*.