



## INFO Data Science Engineering Methods and Tools

### Course Information

Course Title: Data Science Engineering Methods and Tools

Course Number: INFO 6105

Term and Year: Fall 2022

#### CRN:

Course Format: On-Ground

### Instructor Information

Full Name: Pramod Gupta

Email Address: [p.gupta@northeastern.edu](mailto:p.gupta@northeastern.edu)

### Teaching Assistant Information. TBD

Full Name:

Email Address:

Office Hours:

### Course Prerequisites

Graduate level INFO 5100 Minimum Grade of B- or Graduate level CSYE 6200 Minimum Grade of B-

### 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.

## 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.*

*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 Recommended Course Textbooks.

*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.

*An Introduction to Statistical Learning with Applications in R*, Gareth James, et al., Springer, 2013, ISBN-10: 1461471370, ISBN-13: 978-1461471370.

*Time Series Analysis: With Applications in R*, 2nd Edition, Cryer and Chan, Springer, 2010, ISBN-10: 0387759581, ISBN-13: 978-0387759586.

## Course Schedule/Topics Covered.

• Introduction: Concepts in Machine Learning and Predictive Analytics
• Intro to python and Data Pre-processing
• Classification: Nearest Neighbors, Naïve Bayes
• Classification: Decision Tree, CART
• Forecasting: Regression
• Finding Groups of Data: Clustering (K-means and Hierarchical Method)
• Evaluating and Improving Model Performance
• Overfitting and Regularization • Applying Machine Learning Guidance and Practical Issues
• Project Presentation

### Grading Scale.

Assignments: 40%

Project: 60%

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-	

**Attendance/Late Work Policy.** Please insert what is applicable for your class. See sample provided below.

#### Attendance Policy

Students are expected to complete course readings, participate in class discussions or other learning activities during the unit, and complete written assignments for each unit during the time of that unit. It is understood that there might be one week when active participation in ongoing class conversations and learning activities might be delayed. Beyond one week's time, if there is an absence or lateness in participation (1) faculty must be notified in advance; (2) grades will be adjusted accordingly.

## 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 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, 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.

## 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/uhrs>.

## 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>.

### **24/7 Canvas Technical Help**

For immediate technical support for Canvas, call 617-373-4357 or email [help@northeastern.edu](mailto:help@northeastern.edu)

Canvas Faculty Resources: <https://canvas.northeastern.edu/faculty-resources/>

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](mailto: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.