Northeastern University
College of Engineering

INFO 6105 Data Science Engineering Methods and Tools

Course Information
Course Title: Data Science Engineeing Methods and Tools
Course Number: INFO 6105
Term and Year: Spring 2022
Credit Hour: 4
Course Format: On-Ground

Instructor Information
Full Name: Pramod Gupta
Email Address: peegee_2000@hotmail.com

Course Prerequisites
Graduate Level CSYE 6200 Minimum Grade of B- or Undergraduate Level INFO 5100 Minimum Grade of B- or Graduate Level INFO 5100 Minimum Grade of B-

Course Description
Introduces the fundamental techniques for machine learning and data science engineering. Discusses a variety of machine learning algorithms, along with examples of their implementation, evaluation, and best practices. Lays the foundation of how learning models are derived from complex data pipelines, both algorithmically and practically. Topics include supervised learning (parametric/nonparametric algorithms, support vector machines, kernels, neural networks, deep learning) and unsupervised learning (clustering, dimensionality reduction, recommender systems). Based on numerous real-world case studies.

Standard Learning Outcomes

At the conclusion of the course, you will have a basic understanding of machine learning techniques and know how to apply basic machine learning tools in practical situations. Students will complete several assignments and a project using supervised or unsupervised learning algorithms.
- Students should have learned about machine learning techniques,
- Understood how various machine learning algorithms work,
- Apply appropriate tools for various tasks such as prediction, regression , grouping, etc.

Course Detail

Module 1: Introduction

- What is Machine Learning and Data Mining?
- Concepts in Machine Learning an Predictive analytics using simple problems
- What is Predictive Analytics and Machine Learning

- Various Paradigms in Machine Learning
- Steps in developing ML application
- Application of ML
- Introduction to Python and various packages

Module 2: data Preprocessing/Feature Engineering

- Introduction to Data preprocessing and why it is needed
- Various steps/techniques involved.
- Measure of Data Quality
- Major task involved in data preparation
- Dealing with missing values, outliers, data transformation etc.

Module 3: Linear Classifiers

- Linear Regression
- Logistic Regression
- Linear classifiers – common straight forward classifiers with practical applications

Module 4 : Baysian Classifiers

- Histograms classifiers (Naïve Bayes)
- Probability density functions
- Bayesian classifiers
- Class-conditional density, priors and posteriors

Module 5: Classifiers Contd.

- K-nearest neighbor algorithm

Module 6: Decision tree and Ensemble methods

- Decision Tree methods
- Random Forest

Module 7: Unsupervised Learning

- Clustering
- K-means, Hierarchical Methods)
- Expectation maximization algorithm
- Outlier and anomaly detection

Module 8: Feature Selection/Reduction

- Dimension Reduction
- Covariance Matrix
- Feature selection and Principal component analysis
- Regularization

Module 9: Performance Evaluation of Algorithms and Practical Issues

- Evaluating and Improving Model Performance
- Classifier performance evaluation
- Accuracy, sensitivity, specificity, positive predictive value
- Receiver operating characteristic
- Cross validation Training, testing and validation
- Applying Machine Learning Guidance and Practical Issues

Session 10

- Individual/Group Project Presentation

* May cover brief introduction to Neural Networks/Deep learning.

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, 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 HuskyMail 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 Blackboard Technical Help

For immediate technical support for Blackboard, call 617-373-4357 or emailhelp@northeastern.edu

Within Blackboard, open a support case via the red support button on the right side of the screen, click Create Case

myNortheastern, e-mail, and basic technical support
Visit the Information Technology Services (ITS) Support Portal
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, member 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 www.northeastern.edu/titleix for a complete list of reporting options and resources both on- and off-campus.