INFO 6105 – Data Science Engineering Methods and Tools

This course will introduce the students to many concepts, techniques, and algorithms in data mining, machine learning, and statistical modeling including linear regression, classification, logistic regression, regularization, decision trees, random forests, boosting, and feature engineering.

Tools We will use Python to illustrate concepts, analyze data sets, build predictive models, and evaluate the fit of the models.

Instructor: Ebrahim Nasrabadi e.nasrabadi@northeastern.edu

Textbook: An introduction to Statistical Learning with Applications in R, G. James, D. Witten, T. Hastie and R. Tibshirani

Grading:

Assignments	40%
Quiz and Class participation	20%
Midterm Project	10%
Midterm Exam	10%
Final Project	10%
Final Exam	10%

Problem Sets and Case Studies:

- There are 4 problem sets. They are an important part of the learning experience, which is why they are required.
- Problem sets and case studies are to be handed in individually; however, students may discuss the problems with others in the class. *Copying from another students is not permitted.*

No Laptops or Smart Phone Usage in Class: Laptops should not be used in class for anything except taking notes and working on group or individual in class projects. Any other use is not permitted.

"Roughly Schedule"

1	Introduction, Applications, and Concepts
2	Linear Regression
3	Polynomial Regression and Regularization
4	Classification and Logistic Regression
5	Model Evaluation and Cross-validation
6	Model Selection and Regularization
7	In class Lab: Demand forecasting & Loan Application
8	Midterm & Project
9	Regression Trees
10	Classification Trees
11	Random Forests and Boosting
12	Gradient boosting
13	Neural Networks
14	Convolutional neural network
15	In Class Lab: Image Detection
16	Final Exam & Project