

Data Science with Python

MET – CS677

Faculty: Farshid Alizadeh-Shabdiz, PhD, MBA

alizadeh@bu.edu

Office hours: by appointment

Course Description

At the present time, there is a growing need for specialists with background in Python who can apply data science methods to practical problems at their workplace. Working in data science requires an understanding of many interdisciplinary concepts, involves data mining and application of various methods.

The proposed course is designed to fill this need. Students will learn major Python tools, machine learning classifiers and techniques for data analysis. There are weekly assignments and mini projects on topics covered in class. These assignments will help build necessary statistical, visualization and other data science skills for effective use of data science in a variety of applications including finance, time series analysis and recommendation systems. In addition, students will choose a topic for a final project and present it on the last day of class.

The proposed course can be taken by students with not exclusively computer science backgrounds who have basic knowledge of Python.

Books

Required:

“Python for Data Analysis”, by W. McKinney, O’Reilly Publishing, 2017 (2nd edition), ISBN-13: 978-1491957660, purchased from Barnes & Noble

Recommended:

“Python Data Analysis” by Armando Fandango, Packt Publishing, ISBN-13: 978-1787127487

“Python Data Science Handbook” by Jake VanderPlas, O’Reilly Publishing, ISBN-13: 978-1491912058

Courseware

Blackboard Course Notes

Additional materials will be added to “Content (by week)” section.

Class Policies

Weekly programming assignments submitted through blackboard on-line. Late homework is not accepted.

Final projects are submitted through blackboard on-line. Students will present their projects on the last day of class.

The final exam is closed-book and in-class

Academic Conduct Code – “Cheating and plagiarism will not be tolerated in any Metropolitan College course. They will result in no credit for the assignment or examination and may lead to disciplinary actions. Please take the time to review the Student Academic Conduct Code:

Academic conduct code as specified below:

http://www.bu.edu/met/metropolitan_college_people/student/resources/conduct/code.html.

NOTE: [This should not be understood as a discouragement for discussing the material or your particular approach to a problem with other students in the class. On the contrary – you should share your thoughts, questions and solutions

Grading Criteria:

Grades will be computed with following percentages:

- 5% Class participation
- 30% homework
- 20% quizzes
- 15% final project
- 30% final

Tentative Class Syllabus:

The course is divided into 6 modules. Each module will be covered in 2 weeks.

Module	Topic	Reading Due
1	Review of Python, Numpy and data analysis libraries. Intro to statistical machine learning.	Chapter 2 and 3 of the text book.
2	Pandas, Matplotlib & Seaborn, error metrics, model selection trade-offs	Course notes

3	Classification assessment. K nearest neighbor. Logistic Regression. Pandas deep dive.	
4	Data wrangling. Model evaluation. Data scaling.	
5	Clustering algorithms. Kmeans. Hierarchical clustering. DBSCAN. Spectral Clustering. Clustering assessment.	
6	Kmeans cont. DBSCAN variations. BIRCH clustering algorithm. CLIQUE clustering algorithm.	Course notes
7	Maximum Likelihood Est. Generalized linear model. Data cleaning.	Course notes
8	Linear and Quadratic Discriminant Analysis. Bayes rule and Naïve Bayesian Classification.	Course notes
9	Decision trees. Ensemble learning. Random forest classifier.	Course notes
10	Support Vector Machine. Kernel SVM.	Course notes
11	Boosting and bagging algorithms, like AdaBoost and XGBoost.	Course notes
12	Course review, project presentations Course notes and final exam	Course notes