

CSCI381 Data Analytics -- logistics

Course Description:

- Data science has been one of the fastest growing professions recently. The goal of the Data Analytics class is to prepare students with the necessary skill set and understanding to succeed in this area.
- The first part of the course will go over **fundamental concepts** spanning across statistics, cross validation, data visualization, data warehousing and python as data manipulation and model development platform.
- Second part of the course will cover **common machine learning techniques** such as linear and logistics regression, support vector machine, decisions trees and natural language processing.
- Students at the end of the course should be able to **carry on to more advanced studies** on machine learning and **start a career in the data science areas**.

CSCI381 Data Analytics -- logistics

Instructor: Dr. Alex Pang

Email: chiuyan.pang@qc.cuny.edu

Lectures: Mon, Wed (8:00pm – 9:15pm)

Pre-requisites:

- CSCI 313 (Data Structures)
- Math 241 (Prob & Stat)

Teaching Assistant: None

Office hours: Sat, mainly online

Course Objective:

At the end of this course students should

1. have a good overview of the data science professions and modern data analytics platforms.
2. have acquired expertise in using Python as his/her data analysis and model development platform
3. have developed a good analytical mindset in drawing insights on data and making recommendations
4. have understood some of the most common machine learning techniques and feel comfortable in pursuing more advanced skill set in machine learning areas.

CSCI381 Data Analytics -- logistics

Textbook:

Data Analytics Made Accessible:

2019 edition by Anil Maheshwari

Acknowledgement:

I would like to express my special thanks to Dr. Anil Maheshwari for writing such a wonderful textbook in this exciting field as well as his generosity in sharing some of his PowerPoint slides related to his textbook, some of which have been adapted into our course materials

Optional Textbooks (i.e. nice to have, but not necessary for the course):

- Python for Data Analysis by Wes McKinney
- An Introduction to Statistical Learning by Gareth James, Daniela Witten, et al (<http://www-bcf.usc.edu/~gareth/ISL/>)
- Hands-On Machine Learning with Scikit-Learn and Tensor Flow by Aurelien Geron

CSCI381 Data Analytics -- logistics

Section 1: Core Business Intelligence and Data Analytics Concepts

1. Data Science / Data Analytics Overview
2. Probability and Statistics Review
3. Exploratory Data Analysis
4. Data Visualization
5. Common Data Scientist's Toolbox

Section 2: Popular Data Mining Techniques

6. Linear Regression
7. Classification and Logistics Regression
8. Support Vector Machine
9. Decision Trees
10. Clustering
11. Text Mining
12. Big Data

CSCI381 Data Analytics -- logistics

Section 3: More Advanced Techniques and Application

13.Application of Machine Learning

14.Neural Network and Deep Learning

15.Amazon AWS, Microsoft Azure

CSCI381 Data Analytics -- logistics

Grade contribution:

- 40% Homework assignments
- 20% mid-term exam
- 20% final project
- 20% final exam
- 3% extra credits:
 - class participation, end-of-term evaluation and New York open data week competition,
 - absolutely not needed to do well in class

Homework format:

- Python 3 Notebook

Exam format:

- Multiple choices and written short answers

CSCI381 Data Analytics -- logistics

Collaboration Policy:

You are allowed and encouraged to discuss homework. Discussion on Blackboard is encouraged, so everyone can benefit; however, do NOT post or share solutions or parts of solutions. Homework and final project must be done and written up independently.

Academic Integrity Policy:

Absentees are solely responsible for catching-up. Academic dishonesty, such as plagiarism or cheating - taking other people's work with or without their permission in order to get credit for yourself, will be dealt with seriously, including an "F" grade for the course and/or disciplinary action according to the University's policy on academic integrity

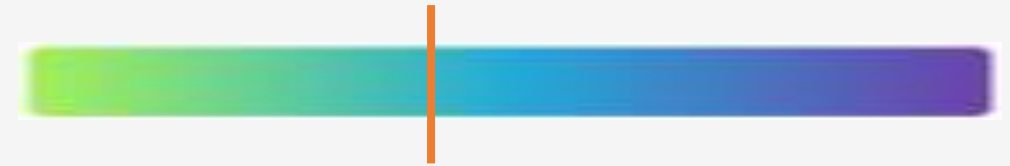
CSCI381 Data Analytics -- logistics

Teaching Style:



Theory

Practical



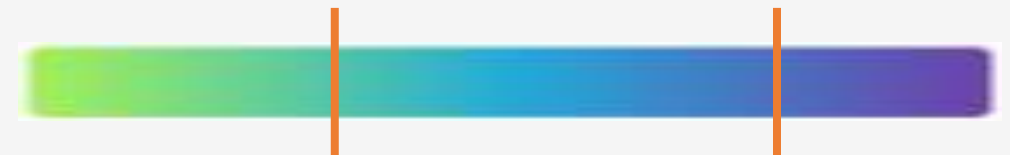
No homework

20 hours per week



Dry
Only me talking

Stand-up comedy
Highly Interactive



follow textbook

No textbook

CSCI381 Data Analytics -- logistics

Weekly Routine:

- Monday:
 - Theory, bi-weekly homework due (if any)
 - Finished reading assignments from the textbook
- Wednesday:
 - Application, homework description,
 - Weekly summary, next week preview, assign reading from the textbook
- Saturday
 - Informal online office hours, will announce before hand, emails
- Will post the PowerPoint after class on Blackboard

CSCI381 Data Analytics -- logistics

Python:

- Lectures as well as homework will be based on Python 3 notebooks
- So you will need to make sure you have a PC or laptop where you can run Python notebooks
- Recommended distribution and installation is Anaconda (<https://www.anaconda.com/>)
- Make sure you are familiar with the syntax as well as the Pandas and NumPy library for data manipulations
- The textbook has a chapter on Python !