SANTA CLARA UNIVERSITY Information Systems & Analytics

Instructor: Denis Vrdoljak
Office: Lucas Hall
Email: dvrdoljak@scu.edu
Class hours: Sat. 8:30AM - 11:15AM // 12:00PM - 2:45PM, in Lucas 309

Office hours: right after class or Wed. 6:00 to 7:00 PM

Teaching Assistant: Robinson Lu blu3@scu.edu , Office hours: by appointment

Course objectives

The objective of this course is to teach the analytical mindset & programming skills relevant to data science. Students will learn the Python programming language, along with a set tools for data science in Python, including the Jupyter (IPython) Notebook, NumPy, Pandas, Seaborn, and Scikit-learn. Students will learn skills that cover the various phases of exploratory data analysis: importing data, cleaning and transforming data, algorithmic thinking, grouping and aggregation, visualization, time series, statistical modeling/prediction and communication of results. The course will utilize data from a wide range of sources and will culminate with a final project and presentation.

Course Material

We will be using a number of resources throughout the class.

- **Camino**. We will use SCU's Camino for all course discussions and to distribute course materials.
- **Online Python Tutorial** (https://www.codecademy.com/learn/python) on codecademy.com. This

  is an excellent, self-guided tutorial on the Python programming language. We will only spend minimal time in class covering the basics of Python; instead you will learn it by completing this online tutorial. You will be required to complete this course by the 3rd week.

- **Python for Data Analysis** https://www.amazon.com/Python-Data-Analysis- Wrangling-IPython/dp/1491957662/ref=dp_ob_title_bk , Wes McKinney, O'Reilly (2nd Edition).

  Grading Policy

- Online Python Tutorial: 10%
- Homework: 20%
- Quizzes: 15% (in-class, no make-up)
- Midterm exam: 20% (in-class, no make-up)
- Group project and presentation: 15%
- Final exam: 20% (in-class, no make-up)

  Important Dates

  Midterm exam: TBD (~week# 6) during class time Final exam: Final exam week (week #10)
  Group project presentations: week before final exam

Software and computers

SANTA CLARA UNIVERSITY Information Systems & Analytics

• This course will involve extensive programming and computer work, both in and out of class. **You are required to bring a laptop to class each time** and have the following software installed by the first class meeting:

- Download Anaconda Distribution: https://repo.anaconda.com/archive/
- Use **Anaconda 3-5.2.0 version (with Python 3.6)** ONLY:
- Direct link for Mac: Anaconda3-5.2.0-MacOSX-x86_64.pkg
- Direct link for Windows: Anaconda3-5.2.0-Windows-x86_64.exe
- Note: Latest Anaconda 2018.12 with Python 3.7 will not work for some packages we will install later

Course Content

The course content is subject to changes based on feedback throughout the class. The course will cover some or all of the topics below.

Part 1: Becoming a data janitor

1. Getting started

- Install software
- The Jupyter notebook environment
- Review Python
- Collecting Twitter data through *tweepy*
- *Numpy*

2. Pandas

- The *Series* class
- The *DataFrame* class
- Data ingestion (csv, text, Excel)

3. Data wrangling

- Data cleaning
- Data transformation
- Data aggregation
- Merging data

4. Data visualization

• *Seaborn*

Part 2: Becoming a data scientist

5. Machine learning: classification

- Introduction to the package *scikit-learn*
- Classification for data exploration using decision trees
- Classification for prediction
- Measuring classification performance

6. Machine learning II: regression

- Regression for data exploration using Lasso and regression trees
- Regression for prediction
- Measuring regression performance

7. Machine learning III: clustering

• Clustering for data exploration

Academic Integrity

Students are expected to uphold the principles of the Academic Integrity for all work in this class:

"Students are expected to do their own work and to cite any sources they use. A student who is guilty of a dishonest act in an examination, paper, or other work required for a course, or who assists others in such an act, may, at the discretion of the instructor, receive a grade of "F" for the course. In addition, a student found guilty of a dishonest act may be subject to sanctions, up to and including dismissal from the University, as a result of the student judicial process as described in the Student Handbook. A student who violates copyright laws, including those covering the copying of software programs, or who knowingly alters official academic records from this or any other institution is subject to similar disciplinary action."

You are encouraged to discuss homework assignments with other students, but you are NOT allowed to turn in the same file. To ensure that you stick to this policy, avoid looking at other people's work.

Disabilities Resources

If you have a disability for which accommodations may be required in this class, please contact Disabilities Resources, Benson 216, http://www.scu.edu/disabilities as soon as possible to discuss your needs and register for accommodations with the University. If you have already arranged

MSIS2802/IDIS3802 Winter 2019 DATA SCIENCE WITH PYTHON accommodations through Disabilities Resources, please discuss them with me during my office hours.

Students who have medical needs related to pregnancy may also be eligible for accommodations.

While I am happy to assist you, I am unable to provide accommodations until I have received verification from Disabilities Resources. The Disabilities Resources office will work with students and faculty to arrange proctored exams for students whose accommodations include double time for exams and/or assisted technology. (Students with approved accommodations of time-and-a-half should talk with me as soon as possible). Disabilities Resources must be contacted in advance to schedule proctored examinations or to arrange other accommodations. The Disabilities Resources office would be grateful for advance notice of at least two weeks. For more information you may contact Disabilities Resources at 408-554-4109.

Accommodations for Pregnancy and Parenting

In alignment with Title IX of the Education Amendments of 1972, and with the California Education Code, Section 66281.7, Santa Clara University provides reasonable accommodations to students who are pregnant, have recently experienced childbirth, and/or have medically related needs. Pregnant and parenting students can often arrange accommodations by working directly with their instructors, supervisors, or departments. Alternatively, a pregnant or parenting student experiencing related medical conditions may request accommodations through Disability Resources.

Discrimination and Sexual Misconduct (Title IX)

Santa Clara University upholds a zero-tolerance policy for discrimination, harassment and sexual misconduct. If you (or someone you know) have experienced discrimination or harassment, including sexual assault, domestic/dating violence, or stalking, I encourage you to tell someone promptly. For more information, please consult the University's Gender-Based Discrimination and Sexual Misconduct Policy at http://bit.ly/2ce1hBb or contact the University's EEO and Title IX Coordinator, Belinda Guthrie, at 408-554-3043, bguthrie@scu.edu. Reports may be submitted online through https://www.scu.edu/osl/report/ or anonymously through Ethicspoint https://www.scu.edu/hr/quick-links/ethicspoint/