CS 495/595 Data Science with Python
Spring 2020

Instructor  M. Zubair
Meeting Time and Place
Thursday 4:20PM - 7:00PM, ECSB 3104

Office Hours
Thursday: 2:30PM - 4:00PM, ECSB 3323

If you can not see me during my office hours, you may also see me at other times by making an appointment. You may send me an email at zubair@cs.odu.edu (email is preferred means of communication). You may also leave messages, papers, etc. for me at the CS Department Office.

Course Description
The objective of this course is to introduce data science using Python programming language. The course will be a hands-on course and requires access to laptops/computers in the class. You can bring your laptop in the class or use one of the machines in the classroom. In the class, lectures will be interleaved with lab exercises. One of the components of this course is a class project where the students will apply the knowledge acquired during the course for a real application. A tentative list of topics is given below.

- Basics of Python programming using Jupyter Notebook
- Use of libraries like Numpy and Scipy, and Data manipulation using Pandas
- Computing basic statistics on data
- Working with different types of data from different application domains
- Plotting and visualization
- Introduction to basic machine learning models

Text and Reading Material
1. Doing Data Science, By: Cathy O'Neil; Rachel Schutt, Publisher: O'Reilly Media, Inc. URL: http://proquest.safaribooksonline.com.proxy.lib.odu.edu/book/databases/9781449363871
3. Deep Learning, By: Ian Goodfellow and Yoshua Bengio and Aaron Courville, Publisher: MIT Press. URL: http://www.deeplearningbook.org
4. As this course is covering new and fast changing content, it also relies heavily on the material available on the Internet. While covering a topic, we will provide the additional Internet pointers.

Prerequisites
Elementary Statistics and Probability; Basic programming (any language): variables, data types and expressions, assignment, control-flow statements, functions, arrays, structs/records; and familiarity with Windows and Unix OS.

Grading
Your grade will be based on a total of 100 points with the following distribution.

- In Class/Take Home Exercises: 20 points
- Take Home Examination: 30 points
- Final Project: 50 points

Schedule for Final Project Presentation: April 18 and April 25 (4pm - 7pm)
The project presentation consists of two parts: presentation, and demonstration of working of the
The final grade will be based on the normalized score.

NOTE Students taking this course at 500 level will be evaluated differently. In particular, I expect much more on a course project from a student at 500 level. If you need clarification, please talk to me.

Project

The course project will enable students to apply their knowledge acquired in the course to develop and implement a data science application. The project should have both the components: exploratory data analysis (EDA) and modeling. You can work alone or in a group of not more than two students on a project. Students are encouraged to work with real data available on sites such as Kaggle and also explore how others are using the data for modeling and EDA. However, please make sure that the significant work on the project should be yours. If you use some parts developed by others, please cite it so that I know what your contribution is and what you have used from others. You can also explore data science competition posted on the Kaggle website. For example:

- New York City Taxi Trip Duration, https://www.kaggle.com/c/nyc-taxi-trip-duration

Course Challenge: You can participate in one of the competitions and if you make it in the top ten, you will be guaranteed a good grade. Please talk to me if you want to pursue this.

You are expected to complete the project in several phases.

Project Kick Start Phase (During 7th week): Finalize your group and discuss/email me the initial project idea for feedback.

Project Finalize Phase (During 8th week): You need to submit your project description. The project description should have the following structure:

- Project Title:
- Project Objective (three or four lines):
- Project Description: One paragraph describing the data sets and the associated data science application. Please make sure you have both the components in your project: EDA and Modeling.

Please submit a plain text file with the above structure. In case, I have issues with your project proposal I will email you.

Project Design Submission Phase (During 10th week): The project design should give implementation details along with a high-level description of the algorithm you will implement. For a high-level description you can use pseudocode or flow charts. The design should clearly identify how you plan to test the application and what data set you will be using for this testing. If you have any questions or concern you can post your question to the instructor in the forum titled: “Problems and Concerns”.

Project Presentation (TBA): The project presentation consists of two parts: presentation, and demonstration of working of the project. The presentation should include an overview of the data science application, approach taken, implementation details, highlighting the EDA and Modeling component of the project and lessons learned.

Academic Honesty and Honor Code

Everything turned in for grading in this course must be your own work. The instructor reserves the right to question a student orally or in writing and to use his evaluation of the student's understanding of the assignment and of the submitted solution as evidence of cheating. Violations will be reported to the Honor Council for consideration for punitive action. All students are expected to abide by the ODU Honor Code.
Accessibility

Old Dominion University is committed to ensuring equal access to all qualified students with disabilities in accordance with the Americans with Disabilities Act. The Office of Educational Accessibility (OEA) is the campus office that works with students who have disabilities to provide and/or arrange reasonable accommodations.

1. If you experience a disability which will impact your ability to access any aspect of my class, please present me with an accommodation letter from OEA so that we can work together to ensure that appropriate accommodations are available to you.
2. If you feel that you will experience barriers to your ability to learn and/or testing in my class but do not have an accommodation letter, please consider scheduling an appointment with OEA to determine if academic accommodations are necessary.

The Office of Educational Accessibility is located at 1021 Student Success Center and their phone number is (757)683-4655. Additional information is available at the OEA website: http://www.odu.edu/educationalaccessibility/

TA/Grader

MAHSA SHARIFI, Email: mshar004@odu.edu
office hours: Monday, 9:am - 12:00pm, Room 3318 (ECSB)

Course Schedule (Tentative)

Please visit the site periodically to view updates.

NOTE: I am including Jupyter Notebooks that I will be covering in the class. Please note these are not meant to replace my lectures.

<table>
<thead>
<tr>
<th>Lect #</th>
<th>Date</th>
<th>Topics (Handouts)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>January 16</td>
<td>Overview resources zip file</td>
<td>Introduction to Jupyter Notebook and Python Basics</td>
</tr>
<tr>
<td>2</td>
<td>January 23</td>
<td>Introduction to Numpy I</td>
<td>Introduction to Numpy I</td>
</tr>
<tr>
<td>3</td>
<td>January 31</td>
<td>Introduction to Numpy II</td>
<td>Introduction to Numpy II</td>
</tr>
<tr>
<td>3</td>
<td>February 6</td>
<td>Basic Statistics I</td>
<td>Basic Statistics</td>
</tr>
</tbody>
</table>