

Course title: Introduction to Big Data and Machine Learning, Fall 2019

Instructor: Dr. R. Paul Mihail, 2119 Nevins Hall, Email: rpmihail@valdosta.edu

Class meeting times and location: Nevins Hall, Room 2115, Tuesday, Thursday from 5:00 PM to 6:15 PM

Office Hours: TR 10:00am-2:00pm and by appointment.

Required Textbooks:

- Bishop, Christopher M. Pattern recognition and machine learning. Springer, 2006.
- Barber, David. Bayesian reasoning and machine learning. Cambridge University Press, 2012.
- Venkat Ankam. Big Data Analytics Paperback. Packt. September 28, 2016.
- Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. Mathematics for Machine Learning. Cambridge University Press, 2019.

Software: The main programming environments will be Python and/or MATLAB.

Datasets:

1. <https://elitedatascience.com/machine-learning-projects-for-beginners>
2. <http://archive.ics.uci.edu/ml/datasets.html>

Course Description: Big data refers to any technology, both storage and compute, that processes very large amounts of data, either in the form of billions of files, or very large files (terabytes to petabytes). In this course, we will dive into Hadoop and Spark, two technologies that are used for distributed storage and compute. Machine learning is a branch of artificial intelligence that focuses on algorithms that learn from experience and adapt to their environments. Machine learning studies data-driven algorithms capable of mimicking, understanding and aiding human and biological information processing tasks. Machine learning methods have been successfully applied to a variety of domains including text, audio, video, imagery, credit card fraud detection, store product placement, operating system scheduling, social networks and many more. In this self-contained course, students will get exposure to theoretical and practical issues involved in the

design of learning systems and big data systems. Students will choose a final project based on their own interests.

Cell phone policy:

Usage of cell phones in the classroom is prohibited, with the exception of emergency calls. Texting or otherwise using your mobile device in the classroom will result in dismissal from the class period.

Learning Outcomes: Students will understand basic programming terminology and techniques. More specifically students will:

1. Describe what is involved in distributed data storage and processing.
2. Describe what Hadoop and Spark clusters are and how to use them.
3. Describe what is involved in learning from data.
4. Develop computer programs that use machine learning algorithms.
5. Describe how to perform evaluation of learning algorithms.
6. Describe how to do model selection.
7. Apply dimensionality reduction methods to data as a preprocessing step.

Course Prerequisites: Students should have a strong programming background, a basic understanding of calculus and exposure to statistics and probability.

Assessment:

The grade for this course will be calculated as follows:

- In-class attendance: 10%
- Programming assignments 70%
- Final Project: 20%

Grades will be assigned according to the following scale:

90-100% = A
80-89.99% = B
70-79.99% = C
60-69.99% = D
Below 60% = F

Exams:

- Midterm exam scheduled during the semester.

What to do if you miss...

- **a lecture** - find out what the material covered was, read the book, borrow someone's notes, find out what any announcements or assignments were. If attendance was taken and you have a documented excuse as described in the attendance policy, contact your professor within one week of your absence.
- **a test** - if you know ahead of time you must miss a test, contact your instructor and make arrangements for an alternate time. If circumstances force you to miss an exam unexpectedly, you **MUST** contact your instructor within a week after the test, in order to have a chance to be allowed to make the exam up.
- **a deadline on an assignment** - see the late policy.

Due dates

- The electronic submission of programs will be done via the course web site. It is not acceptable to email your submission, unless there are technical difficulties and you are instructed to do so.
- Late Policy for Labs - Laboratory programming assignments may not be turned in late, failure to complete the work will result in assignment grades of 0.

Academic Honesty:

Students are expected to do their own work. Cheating is considered a serious offense by the University. Any form of "seeking an unfair academic advantage" is considered cheating.

Withdrawing:

If you decide to leave the class, please do it officially. There is a date on the Academic Calendar past which you are not allowed to drop for academic reasons. We'd much rather give a W grade than an F. Don't just stop coming to class - you **WILL** get an F! Take care of your transcript! All policies associated with this course are subject to revision. Reasonable notification will be provided to students prior to any major changes.

Withdrawal Policy (5 W Policy): Effective Fall 2010, all undergraduate students are limited to five course withdrawal (W) grades for their entire enrollment at Valdosta State University. Once a student has accumulated five W grades, all subsequent withdrawals (whether initiated by the student in BANNER or initiated by the instructor on the proof roll) will be recorded as WF. The grade of WF is calculated as an F for GPA purposes. To get more details about this policy, students are strongly recommended to check the following link:
<http://www.valdosta.edu/academic/WithdrawalPolicy.shtml>

Extra Help: Do not hesitate to come to my office during office hours or by appointment to discuss a homework problem or any aspect of the course. There are also tutors available Monday through Friday, see Mr. Said Fares (office in 1126 Nevins Hall) for more information. There is also the

Student Success Center on campus located on the ground floor of the Langdale Residence Hall. The Student Success Center offers free one-on-one tutoring for core courses, success workshops, etc. You can find more information at <http://www.valdosta.edu/academics/student-success-center/>.

Attendance Policy: Please keep in mind that attendance is extremely important for this course. You are expected to show up for lectures and participate. In case you have to miss class, please make sure you ask for notes or see your professor.

Accommodation for Disabilities: If you have a documented disability that requires academic accommodations, please contact your professor as soon as possible. In order to receive accommodations in this course, you must provide a Letter of Accommodation from the Access Office for Students with Disabilities located in Farver Hall. The phone numbers are 229-245-2498(V/VP) and 229-219-1348(TTY). Accommodations can be made for all parts of the course. We only make special arrangements for class activities after we receive the letter.

Student Opinion of Instruction: At the end of the term, all students will be expected to complete an online Student Opinion of Instruction survey (SOI) that will be available through SmartEvals. Students will receive an email notification through their VSU email address when the SOI is available (generally at least one week before the end of the term). SOI responses are anonymous to instructors/administrators, and they will be able to access results only after they have submitted final grades. Before final grade submission, instructors will not be able to see any responses, but they can see the percentage of students who have or have not completed their SOIs. While instructors will not be able to see student names, an automated system will send a reminder email to those who have yet to complete their SOIs. Students who withdraw or drop a course will also be sent invitations to complete the Dropped Course Survey. Complete information about the SOIs, including how to access the survey, is available on the SOI Procedures webpage: <https://www.valdosta.edu/academics/academic-affairs/sois/welcome.php>

Title IX Statement: Valdosta State University (VSU) is committed to creating a diverse and inclusive work and learning environment free from discrimination and harassment. VSU is dedicated to creating an environment where all campus community members feel valued, respected, and included. Valdosta State University prohibits discrimination on the basis of race, color, ethnicity, national origin, sex (including sexual harassment and sexual violence), sexual orientation, gender identity, religion, age, disability, genetic information, or veteran status, in the University's programs and activities as required by applicable laws and regulations such as Title IX. The individual designated with responsibility for coordination of compliance efforts and receipt of inquiries concerning nondiscrimination policies is the University's Title IX Coordinator: the Director of the Office of Social Equity, titleix@valdosta.edu, 1208 N. Patterson St., Valdosta State University, Valdosta, Georgia 31698, 229-333-5463.