CS 624: DATA ANALYTICS AND BIG DATA

Instructor: Dr. Fengjiao Wang

Spring, 2021

E-mail: f1wang@odu.edu
Web: https://www.cs.odu.edu/~fengjiao/
Instructor office: E&CS 3206
Class Hours: T/R 11:00 am - 12:15 pm
Office Hours: T 12:15 - 1:15 pm and R 1:30 - 2:30 pm or by appointment
Office Location: Zoom

Course Description

Data analytics and data modeling is in great need in the era of big data. The first half is emphasize on learning data analytics tools and learn how to leverage those tools to analyze data and facilitate decision making based on insights from the data. We will study specialized systems and algorithms that have been developed to work with data at scale including parallel database systems, MapReduce and its contemporaries. The second half of the course is close to traditional data mining, which explores models to extract value in the data. We will study various data mining which work with different types of data (static and streaming data) and can be applied to real-world problems in different domains (recommendation, advertising, social network).

Course Objective

Student learning outcomes:

- Demonstrate ability to analyze real-world data by leveraging big data tools
- Able to get insight from the data by performing various queries using in-memory analysis
- Familiar with creating machine learning programs in Python that implement functions and simple algorithms
- Formulate a real-world problem as a machine learning/data mining problem
- Demonstrate ability to build a machine learning model and use appropriate optimization algorithms to obtain a solution for the model
- Use different machine learning methods appropriately to describe a system
- Implement machine learning applications to analyze, predict a real-world system

Textbook and Reference Materials

There is no required textbook. One recommended textbook is:

- Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman et al. 2020. Boot website: http://www.mmds.org

I may assign weekly reading materials from the recommended textbook and or papers.

Final Grades and Grading Policy

The final grade will count all of the assessments with the following proportions:

- Attendance **10%**
- Assignments (Homeworks & Quizzes) **40%**
- Midterm Exam **20%**
- Final Project **30%**

Grading Scale

A	A-	B+	B	B-	C+	C	C-	D	F
93-100	90-92	87-89	84-86	80-83	77-79	74-76	70-73	60-69	<=59

Course Structure and Class Timeline

Tentative course schedule:

week 1: Course Overview

week 2: Relational Database Management Systems (DBMSs)

week 3: Parallel shared-nothing DBMSs Cloud Deployments (Amazon Redshift)

week 4: MapReduce

week 5: MapReduce Today (Hive), Similarity Matching

week 6: In-memory Analytics (Spark & SparkSQL, MLlib)

week 7: In-depth Spark Tutorial

week 8: Advertising on the Web

week 9: Clustering, Invited Talk

week 10: Dimensionality Reduction

week 11: Large-Scale Machine Learning

week 12: Recommendation Systems

week 13: Social-Network Graphs Mining, Invited Talk

week 14: Mining Data Streams

week 15: Final Review and Final project presentation

Contract and Changes: **This is only a tentative course outline. Your instructor reserves the right to make any necessary changes. Any changes will be announced in class.**

Course Policies

Due Dates

Most assignments are maked with an explicit due date, and are due at the end of that day (11:59:59PM, ET). You will find these dates in course website.

Late submissions for homework and project will be accepted, at a 10% per day penalty, up until one week after due date. Late submissions will not be accepted once the scheduled starting time of the final exam has begun. Late submissions of quizzes and exams are not accepted.

> Exceptions to these dates will be made only in situations of unusual and unforseeable circumstances beyond the student's control.

"I've fallen behind and can't catch up", "I'm having a busier semester than I expected", or "I registered for too many classes this semester" are not grounds for an extension.

Academic Honesty

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 Office of Student Conduct Academic Integrity for consideration for possible punitive action.

Students who contribute to violations by sharing their code/designs with others may be subject to the same penalties.

Students are expected to use standard Unix protection mechanisms ('chmod') to keep their assignments from being read by their classmates. Failure to do so will result in grade penalties, at the very least.

This policy is *not* intended to prevent students from providing legitimate assistance to one another. Students are encouraged to seek/provide one another aid in learning to use the operating system, in issues pertaining to the programming language, or to general issues relating to the course subject matter.

> Students should avoid, however, explicit discussion of approaches to solving a particular assignment, and under no circumstances should students show one another their code for an ongoing assignment, nor discuss such code in detail.

Educational Accessibility

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

- If you experience a disability which will impact your ability to access any aspect of the course, present me with an accommodation letter from OEA so that we can work together to ensure that appropriate accommodations are available to you.
- If you feel that you will experience barriers to your ability to learn and/or complete examinations in the course but do not have an accommodation letter, 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/).