

DATA 301 – Data Science Methodology

Instructor: Keith Perkins

Office: SERC 2275

Office Hours: W 11-12PM
W 1-3 PM
TTh 12:15PM-1:30PM

E-mail: keith.perkins@cnu.edu

Purpose: This course introduces modern statistical and machine learning techniques and demonstrates their application on real world datasets. Topics include advanced clustering algorithms, tree-based data analysis including random forest and gradient boosted trees, and neural network-based systems. Students will use these algorithms to solve real world problems. This is a projects-based course.

Prerequisite: Data 201 with a grade of C- or higher.

Notes and Assignments:

Course material will come from lecture notes, scholar, and the web.

Required Text: None.

Suggested Texts: *Python for Data Analysis*, Wes McKinney – for pandas

Machine Learning with Python Cookbook, Chris Albon – common task recipes

Exams

There will be 2 exams during the semester and 1 final exam at the end of the semester. The final exam is comprehensive.

Projects

There will be up to 5 projects assigned throughout the semester. Due dates will be rigorously observed. **Projects turned in up to 1 week late will be penalized 50%. Projects later than 1 week will not be accepted.**

Make Ups

I will not give make-up exams or accept any late projects except in cases of documented illness or valid justification.

Grading Policy:

Final grades will be based on the following weighting distribution.

Projects	10%
Midterm1	30%
Midterm2	30%
Final Exam.....	30%

Tentative weekly content and assignment calendar

Week	Topics	Test	Project Assign	Project Due
1	Introduction, Tools, Project workflow		Project1	
2	Data loading, cleaning, EDA (review from DATA 201)			
3	“			
4	Clustering algorithms		Project2	Project1
5	“			
6	Splitting a dataset (train, test, val) Data leakage Data imbalance Normalizing	test		
7	Decision Trees, Random Forest Bias/Variance trade off Under/over fitting Correlation and Covariance		Project3	Project2
8	“			
9	Gradient Boosted Trees			
10	Explainability or why a model made a decision (Permutation testing, Variable Importance)	Test	Project4	Project3
11	Neural Networks			
12	Recommender Systems		Project5	Project4
13	Topics			
14	“			Project5
15		Final		
Project List (Complete project description and requirements will be available on scholar)				
15 pts - Project 1: Data Cleaning and display				
20 pts - Project 2: Clustering				
20 pts - Project 3: RF or Boosted Trees				
20 pts - Project 4: Tabular data analysis using multiple algorithms				
25 pts - Project 5: TBD				

Numerical Grading Scale

A	94-100
A-	90-93
B+	87-89
B	84-86
B-	80-83
C+	77-79
C	74-76
C-	70-73
D+	67-69
D	64-66
D-	60-63
F	< 60

Class Conduct:

Treat others in the class with respect. Please feel free to ask questions. Please arrive to class on time. Please turn off cell phones.

Academic Honesty:

Academic honesty is fundamental to the mission of a university. Cheating, plagiarism and any other violations of the Honor Code will result in a failing grade on the assignment or the course. (depending on the severity of the offense). Violations will also be reported to the Center for Honor Enrichment and Community Standards (CHECS).

The Christopher Newport University Honor Code:

“On my honor, I will maintain the highest possible standards of honesty, integrity and personal responsibility. This means I will not lie, cheat or steal, and as a member of this academic community, I am committed to creating an environment of respect and mutual trust.”

University Statement on Diversity and Inclusion:

The Christopher Newport University community engages and respects different viewpoints, understands the cultural and structural context in which those viewpoints emerge, and questions the development of our own perspectives and values, as these are among the fundamental tenets of a liberal arts education.

Accordingly, we affirm our commitment to a campus culture that embraces the full spectrum of human attributes, perspectives, and disciplines, and offers every member of the University the opportunity to become their best self.

Understanding and respecting differences can best develop in a community where members learn, live, work, and serve among individuals with diverse worldviews, identities, and values. We are dedicated to upholding the dignity and worth of all members of this academic community such that all may engage effectively and compassionately in a pluralistic society.

If you have specific questions, suggestions, or concerns regarding inclusion on campus, please contact belonging@cnu.edu.

Disabilities/Accessibility:

In order for a student to receive an accommodation due to a disability, that disability must be on record in the Office of Student Affairs, 3rd Floor, David Student Union (DSU). If you have a diagnosed disability, please contact Jacquelyn Barnes, Director of Accessibility & Care Team Support in Student Affairs (594-7160) or email careteam@cnu.edu to discuss your needs. You may visit <https://cnu.edu/life/acts/accessibility/> to request accommodations or view additional information.

In order to provide access to educational programs in the best manner possible, students with documented disabilities should notify the instructor at least seven days prior to the point at which they require an accommodation (the first day of class is recommended), in private, if accommodation is needed. The instructor will provide students with disabilities with the reasonable accommodations approved and directed by the Office of Student Affairs. Work completed before the student notifies the instructor of his/her disability may be counted toward the final grade at the sole discretion of the instructor.

Success:

I want you to succeed in this course and at Christopher Newport. I encourage you to contact me during office hours or schedule an appointment to discuss course content or ask questions. If I become concerned about your course performance, attendance, engagement, or well-being, I will contact you first. I also may submit a referral through our Captains Care Program. The referral will be received by the Center for Student Success as well as other departments when appropriate. If you are an athlete, the Assistant Director of Student-Athlete Success Services will be notified. Someone from the Center for Student Success will contact you to set up a plan for success and connect you with the appropriate resources. Please remember that this is a means for me to support you and help foster your success at Christopher Newport.

Academic Support:

The Center for Student Success (CSS) provides numerous free resources to help students succeed in their courses. These resources include individual peer tutoring and group/test review sessions. Student Success Coordinators and Student-Athlete Academic Success Coaches can meet with you regularly to discuss your plan for getting your coursework done while balancing your extracurricular activities. They can help you to develop individualized study skills, create time management plans, prepare for exams and papers, etc. The Center also houses the Alice F. Randall Writing Center. Writing consultants can help you at any stage of the writing process, from invention to development of ideas, to polishing a final draft.

You may contact the Center for Student Success to request a tutor, meet with a writing consultant, or make an appointment to talk with a staff member about study skills and strategies. The Center is located in Christopher Newport Hall, first floor, room 124. You may email studentsuccess@cnu.edu, call (757) 594-7684, or visit <https://my.cnu.edu/studentsuccess/>.

Public Health

The university will provide guidance on public health issues and students will be expected to comply with university protocols.

Course Materials:

All content created and assembled by the faculty member and used in this course is to be considered intellectual property owned by the faculty member and Christopher Newport University. It is provided solely for the private use of the students currently enrolled in this course. To ensure the free and open discussion of ideas, students may not make available any of the original course content, including but not limited to lectures, discussions, videos, handouts, and/or activities, to anyone not currently enrolled in the course without the advance written permission of the instructor. This means that students may not record, download, screenshot, or in any way copy original course material for the purpose of distribution beyond this course. A violation may be considered theft. It is the student's responsibility to protect course material when accessing it outside of the physical classroom space.