DATA 301 – Data Science Methodology

Instructor: Keith Perkins

Office: Luter 354
Office Hours: TTh 12:20-1:25, WF 10-12

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 …….....40%
Midterm1 ..……..….20%
Midterm2 ..……...….20%
Final Exam……..…..20%

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 Covarience		Project3	Project2
8	“			
9	Gradient Boosted Trees			
10	Explainability or why a model made a decision (Permutation testing, Variable Importence)	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

Honor Code

The Honor Code will be strictly observed. All work must be done individually. Cheating on any work will result in either a a score of zero or an F for the course, and/or the filing of a case in the CNU honor court. Violation of the honor code may result in dismissal from the University.

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 and beepers.

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 diversity on campus please contact Diversity.Inclusion@CNU.edu

Disabilities:

In order for a student to receive an accommodation for a disability, that disability must be on record in the Office of Student Affairs, 3rd Floor, David Student Union (DSU). If you believe that you have a disability, please contact Jacquelyn Barnes, Student Disability Support Specialist in Student Affairs (594-7160) to discuss your needs.

Students with documented disabilities are to 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 to schedule an appointment to discuss course content or to answer questions you have. During the Coronavirus pandemic, our conversations may need to be via electronic means. 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 Academic Success as well as other departments when appropriate (Counseling Services, Office of Student Engagement). If you are an athlete, the Athletic Academic Support Coordinator will be notified. Someone will contact you to help determine what will help you succeed. 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 Academic Success offers free tutoring assistance for Christopher Newport students in several academic areas. Center staff offer individual assistance and/or workshops on various study strategies to help you perform your best in your courses. 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. The Center is not a proofreading service, but consultants can help you to recognize and find grammar and punctuation errors in your work as well as provide assistance with global tasks. Contact them as early in the writing process as you can!

You may contact the Center for Academic Success to request a tutor, confer with a writing consultant, obtain a schedule of workshops, 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 123. You may email academicsuccess@cnu.edu or call (757) 594-7684.

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.