Data Science 3001: Foundations of Machine Learning

Instructor: Navya Annapareddy
Contact: navya@virginia.edu
Office Hours: Mondays from 10 - 11am (in person, First Floor Hub), Thursdays 5 - 6pm (Online)

Course Description: During DS 3001, you will learn foundational knowledge in each of the four high level domain areas of data science (Value, Design, Analytics, Systems). This includes an emphasis on ethical issues surrounding the field of data science and how these issues originate and extend into society more broadly." This course introduces machine learning and using data and algorithms to create predictive models. We will work with data from medicine, criminal justice, business, biology, geology, and a variety of other sources. This course will equip you with the foundational models and principles of building reliable and robust predictive algorithms. In the process, you will explore the practical and ethical considerations of creating and scaling these tools.

Prerequisites: Student must be in the Data Science minor, have completed one of the DS Minor prerequisite courses (STAT 1100, STAT 1120, STAT 2020, STAT 1601, STAT 2120, APMA 3100, APMA 3110), and must have also completed one of the following: CS 1110, CS 1111, CS 1112, CS 2110, DS 1002, STAT 1601, or STAT 1602.

Course Materials:
We will use 2 texts throughout this course. Selected readings will be available on Canvas.

1. Sebastian Raschka; Yuxi (Hayden) Liu; Vahid Mirjalili; Dmytro Dzhulgakov, Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python , Packt Publishing, 2022.
2. Kuhn, Max & Johnson, Kjell. (2013). Applied Predictive Modeling. 10.1007/978-1-4614-6849-3.

Additional course and assignment materials will be available through GitHub at https://github.com/ncanna/DS-3001.

Learning Objectives:

- Gain experience working in teams to solve data driven problems
- Be able to describe the field of machine learning
- Understand computational and mathematical foundations of the methods presented
- Articulate the advantages and disadvantages of selected ML approaches
- Be able to select appropriate ML models given problems and data types
- Understand the importance of and methods for evaluating ML models
- Gain the foundational knowledge necessary to succeed in future machine learning coursework

Grading Breakdown:
Quizzes (15%) – Each module will feature 1 or 2 short summary quizzes (5 in total). They will be open-note and auto-graded by Canvas. You are allowed unlimited chances to complete the quiz and as such, feedback for specific questions will be hidden. Quizzes should be completed individually.

Labs (60%) – We will have 4 in-class lab assignments throughout the class. Labs are designed to allow you to practice the skills being presented in class. While they will be submitted individually, you are encouraged to work with your peers to the best of your interest. Labs will be submitted on Canvas and while individual submission requirements will differ for each week, you will typically need to submit:

- A publishable markdown or Python notebook document in PDF or HTML format
- Raw code folder (.zip) or link (ie: GitHub)

Final Presentation and Report (25%) – The course will culminate in a final project that will involve working with a dataset of your choice, giving a presentation, and submitting well-annotated code to include summary information in report form. This is an open-ended project designed to allow small groups and individuals to choose a topic of personal interest from the semester and perform exploratory data analysis (EDA) and predictive modelling on. The project will have intermediary assignments embedded within lab sessions. Graded deliverables will be released alongside an evaluation rubric.

- **Data Description (Pass/Fail) -** 1 page description of group member(s), dataset source, size, features, and potential considerations when processing or cleaning data.
- **Pre-Analysis Plan (Pass/Fail) -** 2 page summary of the intended data features and model plan. Plans should include at least 2 EDA visualizations and 2 supervised learning components.
- **Final Presentation (10%) -** Groups will give 15 minute presentations of their findings and analysis. Presentations should include an overview of the dataset, key methods and results, and a discussion of the implications of results.
- **Final Report (15%) -** 5 page collaborative report and code due after presentation during the Finals period. The report will include a description of data preparation, substantial interpretation of results, and discussion of analysis with its ethical implications.

Course Schedule:

Weekly Schedule for DS 3001: Foundations of Machine Learning				
Week		Lecture	Lab Session	Quiz
Module 1: Python for Machine Learning				
Week 1	01/13 - 01/17	Syllabus Review	Environment Setup	
Week 2	01/20 - 01/24	Python for Machine Learning	**Lab 1: Pandas and Data Sourcing**	
Week 3	01/27 - 01/31	APIs and Web Scraping		**Quiz 1**
Module 2: Introduction to Machine Learning Concepts				

Week 4	02/03 - 02/07	Data Preparation	Final Project Group Selections	
Week 5	02/10 - 02/14	Predictive Modeling (Regression)	**Data Description Assignment**	
Week 6	02/17 - 02/21	Predictive Modeling (Regression II)	Final Project Work Time	**Quiz 2**
Week 7	02/24 - 02/28	K-Nearest Neighbors	**Lab 2: Evaluation Lab**	
Week 8	03/03 - 03/07	Sampling and Cross-Validation		**Quiz 3**
Spring Break: 03/08 - 03/16				
Module 3: Decision Trees				
Week 9	03/17 - 03/21	Decision Trees	**Lab 3: Decision Trees**	
Week 10	03/24 - 03/28	Ensemble Methods		**Quiz 4**
Module 4: Unsupervised Machine Learning				
Week 11	03/31 - 04/04	K-Means Clustering	**Lab 4: Unsupervised ML Pipelines**	
Week 12	04/07 - 04/11	Principal Component Analysis		**Quiz 5**
Module 5: Course Wrap-Up				
Week 13	04/14 - 04/18	**Final Presentations**		
Week 14	04/21 - 04/25			

Week 15	04/28	LDOC Lecture
Final Report Due May 7th 11:59pm		

Grading Policy:
Courses carrying a Data Science subject area use the following grading system: A, A-; B+, B, B-; C+, C, C-; D+, D, D-; F. The symbol W is used when a student officially drops a course before its completion or if the student withdraws from an academic program of the University. Your final letter grade will be determined by the following scale. All grades are considered final 5 days after the end of the final examinations period.

Grade	% Lower Bound	% Upper Bound
A+	98.0	100.0
A	97.999	93.0
A-	92.999	90.0
B+	89.999	87.0
B	86.999	83.0
B-	82.999	80.0
C+	79.999	77.0
C	76.999	73.0
C-	72.999	70.0
D+	69.999	67.0
D	66.999	63.0
D-	62.999	60.0

Late Policy: Life happens. While graded assignments will typically have dedicated work time within class, I will accept extension requests before the due date of an assignment with no grade penalty. Late assignments that do not have a reason for being late will be docked a half letter grade per day late. All assignments, regardless of if a late submission was approved, must be submitted by the last day of classes, or else they will be graded as a 0.

Generative AI: The use of generative text (ie: ChatGPT) and generative code tools (ie: CoPilot) are allowed in this class, with the following caveats. **Any lines of code contributed to by generative AI must be cited**. Generative AI should not be used for writing assignments, but may be used in editing and proofreading capacities. Generative AI can be used for brainstorming and receiving feedback. Code or writing suspected of being AI generated to a point of concern may be flagged and followed up.

Honor Policy: All work should be pledged in the spirit of the Honor System at the University of Virginia. The instructor will indicate which assignments and activities are to be done individually and which permit collaboration. The following pledge should be written out at the end of all quizzes, examinations, individual assignments, and papers: "I pledge that I have neither given nor received help on this assignment." The pledge must be signed (name written out on digital assignments). For

more information, visit www.virginia.edu/honor. All suspected violations of the Honor or Generative AI policies will be forwarded to the Honor Committee.

Compute: You do not need an expensive computer for this course. Rivanna HPC will be available for students enrolled in the course, and Google Colab is an online service accessible to anyone. All the content will be hosted on GitHub and all students will have equitable access to computational resources.

Accommodations: The University of Virginia accommodates students with disabilities. Any SCPS student with a disability who needs accommodation (e.g., in arrangements for seating, extended time for examinations, or note-taking, etc.), should contact the Student Disability Access Center (SDAC) and provide them with appropriate medical or psychological documentation of his/her condition. Once accommodations are approved, just follow up with me concerning any logistics and implementation of accommodations. Please try to make accommodations for test-taking at least 14 business days in advance of the date of the test(s). Students with disabilities are encouraged to contact the SDAC, who maintains the following contact methods:

- Voice: 434-243-5180
- Video Phone: 434-465-6579
- Fax: 434-243-5188

Further policies and statements are available at www.virginia.edu/studenthealth/sdac/sdac.html

Diversity and Inclusion: My expectation is that we all contribute to an inclusive and respectful classroom culture that reflects the School of Data Science's commitment to being a space in which you can find true belonging and a sense of shared community. The diversity (referring to the multiple ways that we identify ourselves, including but not limited to race, color, national origin, language, sex, disability, age, sexual orientation, gender identity, religion, creed, ancestry, belief, veteran status, or genetic information) of our classroom is a strength. You are expected to respectfully embrace the opportunity to engage, collaborate, and learn with/from a diverse team of classmates. If you find yourself in need of additional support, please consider the following resources:

- SDS Associate Dean for DEI, Siri Russell: ssr5v@virginia.edu
- UVA Just Report It
- Student Disability Access Center: 434-243-5180