CS 404/504 – Special Topics: Python Programming for Data Science

Credit Hours: 3

Instructor: Alex Vakanski, vakanski@uidaho.edu

Office Location: TAB 311, Idaho Falls Center

Semester: Fall 2022 (August 22 – December 16)

Delivery Methods: Live classroom meetings, or Virtual (Zoom) meetings

Course Description

With the increased use of data science projects for improving various functions and operations across organizations, the tools for managing such projects have matured as well. This course introduces students to Python tools and libraries that are commonly used by organizations for management of the different phases in the life cycle of data science projects. The content is divided into four main themes. The first theme reviews the basics of Python programming and extends it with advanced concepts. The second theme focuses on data engineering, and covers Python tools for data collection and exploration. The next theme overviews model engineering, and includes model design, training, testing, optimization, and packaging. The last theme introduces Data Science Operations (DSOps), and covers techniques for model serving, performance monitoring, diagnosis, and reproducibility of data science projects deployed in production.

The course will provide hands-on Python programming experience for data science workflow management. Each class session will begin with a lecture presentation, and it will continue with reviewing code implementations of studied techniques, libraries, and concepts. Additional work is required for graduate credit.

Learning Outcomes

Upon the completion of the course, the students should demonstrate the ability to:

1. Understand and describe commonly used Python frameworks for life cycle management of data science projects.
2. Apply advanced Python tools for data collection, analysis, and visualization.
3. Design, validate, and justify the selection of data science models using statistical approaches, data mining, and machine learning methods.
4. Implement algorithms for image and natural language processing using Python-based frameworks.
5. Understand the main characteristics of existing Python libraries for deployment, continuous integration, and monitoring of data science projects.
6. Deploy data science models on cloud servers and edge devices.

Prerequisites
CS 212 Practical Python OR CS 477/577 Python for Machine Learning OR Instructor Permission

The course requires having basic programming skills in Python. Knowledge in data science approaches is recommended, but it is not required.

Textbooks
1. Joel Grus, "*Data Science from Scratch: First Principles with Python*," 2nd Edition, O'Reilly Media, 2019, ISBN: 9781492041139.
2. Chip Huyen, "*Designing Machine Learning Systems*," O'Reilly Media, 2022, ISBN: 9781098107963.

Grading
Student assessment will be based on 3 quizzes (worth 30 marks), 4 homework assignments (worth 60 pts), and class participation and engagement (worth 10 marks)**.**

Assessment Component	*Marks*
Quizzes (x3)	30
Assignments (x4)	60
Class participation	10
Total	*100*

Tentative Schedule
Week 1: Introduction
A Short History and Current State of Artificial Intelligence
Python Basics Review: Data Types

Week 2: Python Basics Review (cont'd)
Statements, Files
Iterators, Generators, Functions and Scope

Week 3: OOP
Object-Oriented Programming
Decorators

Week 4: Python Functional Programming
Higher-order Functions, Closures
Meta-programming
Context Managers
Descriptors

Week 5: Data Engineering Pipelines: Data Collection
Scrapping the Web
NumPy for Array Operations

Week 6: Data Engineering Pipelines: Data Exploration
Data Manipulation with Pandas
Data Visualization with Matplotlib

Week 7: Data Engineering Pipelines: Feature Extraction
Feature Selection
Feature Engineering

Week 8: Data Engineering Pipelines: Databases and SQL
Query Tables, Groups, Ordering
Subqueries, Optimization

Week 9: Model Engineering Pipelines: Model Design
Scikit-Learn Library for Data Science
Classification, Regression
Image Classification with TensorFlow

Week 10: Model Engineering Pipelines: Model Design
Natural Language Processing with PyTorch
Large Language Models
Time Series Analysis and Forecasting

Week 11: Model Engineering Pipelines: Model Optimization
Modular Code
Model Selection, Fine-tuning, AutoML
Multi-threading and Multi-processing

Week 12: Model Engineering Pipelines: Model Packaging
Configuring Data Science Projects
Model Serialization Formats
Publishing Python Packages
Git and Version Control

Week 13: Deployment Pipelines: Model Serving
DSOps Tools
Model Serving in a Production Environment

Week 14: Deployment Pipelines: Performance Monitoring
Tools for Monitoring Data Science Projects
Failure Diagnosis, Data Distribution Shifts

Week 15: Deployment Pipelines: Reproducibility
Reproducible Data Science Projects

Docker Containers, Kubernetes
Virtual Environments

Week 16: Deployment Pipelines: Deployment to Mobile Devices and the Web
Deploying Data Science Projects as Mobile Applications
Deploying Data Science Projects to a Web Server with Flask

Academic Integrity
Students are expected to adhere to the highest academic standards of honesty and integrity. At UI, we assume students will do their own work. Plagiarism—passing off someone else's work as your own, without citing the source—should not be tolerated. This includes direct copying, rephrasing, and summarizing, as well as taking someone else's idea and putting it in different words. The best avenue for avoiding plagiarism issues is to fully cite all sources used for preparing assignments, texts, exams.

Learning Civility
In any environment in which people gather to learn, it is essential that all members feel as free and safe as possible in their participation. To this end, it is expected that everyone in this course will be treated with mutual respect and civility, with an understanding that all of us (students, instructors, professors, guests, and teaching assistants) will be respectful and civil to one another in discussion, in action, in teaching, and in learning.

Should you feel our classroom interactions do not reflect an environment of civility and respect, you are encouraged to meet with your instructor during office hours to discuss your concern. Additional resources for expression of concern or requesting support include the Dean of Students office and staff (208-885--6757), the UofI Counseling & Testing Center's confidential services (208-885-6716), the UofI Office of Equity and Diversity (208-885-2468), or the Office of Civil Rights and Investigations (208-885-4285).

Center for Disability Access & Resources (CDAR)
University of Idaho is committed to ensuring an accessible learning environment where course or instructional content are usable by all students and faculty. If you believe that you require disability-related academic adjustments for this class (including pregnancy-related disabilities), please contact Center for Disability Access and Resources (CDAR) to discuss eligibility. A current accommodation letter from CDAR is required before any modifications, above and beyond what is otherwise available for all other students in this class will be provided. Please be advised that disability-related academic adjustments are not retroactive. CDAR is located at the Bruce Pitman Building, Suite 127. Phone is 208-885-6307 and e-mail is cdar@uidaho.edu. For a complete listing of services and current business hours visit https://www.uidaho.edu/current-students/cdar.

Inclusivity Statement
As a professor/course instructor at the University of Idaho, I acknowledge the importance of diversity and inclusion and how these attributes contribute to the promotion of a positive educational experience. It is my intent to facilitate a healthy, productive, and safe learning

environment where diverse thoughts, perspectives, and experiences are welcomed, and individuals' identities (including, but not limited to: race, sex, class, sexual orientation, gender identity, ability, religious beliefs, etc.) are valued and honored. I recognize that as an educator, it is my responsibility to take the initiative to continually learn about diverse perspectives and identities; therefore, if at any point during the course, you feel uncomfortable or concerned, I am more than willing to discuss suggestions, feedback, and anything else that might improve the general effectiveness of this course.

Healthy Vandals Policies
Please visit the University of Idaho COVID-19 webpage often for the most up-to-date information about the UofI's response to Covid-19.

Vandal Food Pantry
The Vandal Food Pantry is a free resource stocked weekly with food, grocery bags, and various hygiene items. Its eight locations across campus are accessible during building hours and open to all. Please take what you need.

Green Dot Safety Program
What's Your Green Dot? It's up to all of us to make a safer campus. Vandal Green Dot is a program that helps students learn about the power of the bystander, how to recognize potentially risky situations, and realistic ways to intervene. Together we can bring down the number of people being hurt by interpersonal violence on our campus. No one has to do everything, but everyone has to do something! Learn more and get involved by visiting UI's Green Dot Safety Program or emailing greendot@uidaho.edu.

Help and Resources
Student Resources
The University of Idaho provides student support to ensure a successful learning experience.

- Student Resources Webpage
- SI-PASS (Peer Assisted Study Sessions) SI-PASS provides regularly scheduled, peer-led study sessions for difficult courses.

Library Help
The UofI Library website has many databases that will help you find relevant and reliable books, articles, images, and more. Don't hesitate to contact a librarian for research assistance.

- UofI Library Website
- Help - Reference Services
- Help for Distance Ed Students

Technology Help
The UofI Student Technology Center provides many technology related services to students.

- PHONE: 208-885-HELP (208-885-4357)
- Technology Help Email: support@uidaho.edu
- Technology Help Website

Writing Support
The UofI Writing Center provides one-on-one assistance to student writers and other members of the campus community.

- PHONE: 208-885-6644
- Writing Center Email: writingcenter@uidaho.edu
- Writing Center Website

UofI Moscow Land Acknowledgement
UofI Moscow is located on the homelands of the Nimiipuu (Nez Perce), Palus (Palouse) and Schitsu'umsh (Coeur d'Alene) tribes. We extend gratitude to the indigenous people that call this place home, since time immemorial. UofI recognizes that it is our academic responsibility to build relationships with the indigenous people to ensure integrity of tribal voices.