CS 4622/5622 – Applied Data Science with Python

Semester: Fall 2025 (August 25 – December 19)

Credit Hours: 3

Instructor: Alex Vakanski, vakanski@uidaho.edu

Office Location: TAB 311, Idaho Falls Center

Course Delivery Methods:

- Virtual meetings (live meetings, students participate through Zoom)
- Classroom (live meetings, video link from Idaho Falls)
- Online (recorded Zoom videos of lectures available to students to watch after the classes)

Course Description

The course introduces students to Python tools and libraries commonly used by organizations for managing the phases in the life cycle of Data Science projects. The content is divided into four main themes. The first theme reviews the fundamentals of Python programming. The second theme focuses on data engineering and explores Python tools for data collection, exploration, and visualization. The next theme covers model engineering and includes topics related to model design, selection, and evaluation for image processing, natural language processing, and time series analysis. This theme also introduces recent advances in large language models, multi-modal models, and agentic AI systems. The last theme focuses on Data Science Operations (DSOps) and encompasses techniques for model serving, performance monitoring, diagnosis, and reproducibility of data science projects deployed in production. Throughout the course, students will gain hands-on experience with various Python libraries for Data Science workflow management. Additional work is required for graduate credit.

Learning Outcomes

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

1. Attain proficiency with commonly used Python frameworks for managing the life cycle of Data Science projects.
2. Develop pipelines for integrating data from multiple sources, designing predictive models, and deploying the models.
3. Apply Python tools for data collection, analysis, and visualization, such as NumPy, Pandas, Matplotlib, and Seaborn, to real-world datasets.
4. Implement machine learning algorithms for image processing, natural language processing, and time series analysis using Python-based frameworks, such as Scikit-Learn, Keras, TensorFlow, and PyTorch.
5. Understand the principles of model selection and evaluation, including hyperparameter tuning, cross-validation, and regularization.
6. Design and implement advanced AI systems using large language models, vision-language models, and agentic AI integration.

7. Understand the primary characteristics of current Python libraries for deployment, continuous integration, and monitoring of Data Science projects.
8. Deploy Data Science projects as web applications using Flask, FastAPI, and Django, and to cloud servers using Microsoft's Azure platform.

Prerequisites
The course requires basic programming skills in Python. Prior knowledge of data science methods is beneficial but not required.

Textbooks
There are no required textbooks for this course.

Grading
Student assessment will be based on 6 homework assignments (worth 45 pts), 6 quizzes (worth 45 marks), and class participation and engagement (worth 10 marks)**.**

Assessment Component	*Marks*
Assignments (x6)	45
Quizzes (x6)	45
Class participation	10
Total	*100*

Tentative Schedule

Date	**Topics, Tests, Assignments**
Week 1: Aug. 25	**Introduction** **Lecture 1**: **A Short History and Current State of Artificial Intelligence** **Python Review** **Lecture 2**: **Python Data Types** – Numbers, Strings, Lists, Dictionaries, Tuples, Sets, Other Types **Tutorial 1:** Jupyter Notebooks
Week 2: Sep. 1	**Python Review** **Lecture 3: Statements, Files** – Conditional statements, loop statements – File handling, opening, writing, appending, iterating through a file, storing objects in files **Lecture 4: Functions, Iterators** – Function definition, function call, argument passing, namespace and scope **Tutorial 2:** Python IDEs, Visual Studio Code
Week 3: Sep. 8	**Python Review**

	Lecture 5: Object-Oriented Programming, Modules, Packages – Class definition, attributes, methods, inheritance, polymorphism, special methods – Creating and using modules, module namespaces, organizing modules into packages **Data Engineering Pipelines** **Lecture 6: NumPy for Array Operations** – Array creation, indexing, broadcasting and vectorization, reshaping arrays, linear algebra with NumPy **Tutorial 3:** Terminal and Command Line Due: Homework 1
Week 4: Sep. 15	**Data Engineering Pipelines** **Lecture 7: Data Manipulation with Pandas** – Importing data, indexing and slicing, merging, sorting, exporting data, dealing with missing values **Lecture 8: Data Visualization with Matplotlib** – State-based approach, customizing plot appearance, saving figures, subplots, plotting functions, object-oriented approach **Tutorial 4:** Virtual Environments **Quiz 1**
Week 5: Sep. 22	**Data Engineering Pipelines** **Lecture 9: Data Visualization with Seaborn** – Creating statistical plots, visualizing relationships between variables, customizing plot styles **Lecture 10: Databases and SQL** – Intro to SQLite, creating databases, querying, sorting, filtering, joining tables, conditional expressions, grouping data, subqueries **Tutorial 5:** Google Colab Due: Homework 2
Week 6: Sep. 29	**Data Engineering Pipelines** **Lecture 11: Data Exploration and Preprocessing** – Exploratory data analysis, preprocessing numerical data, preprocessing categorical data **Model Engineering Pipelines** **Lecture 12: Scikit-Learn Library for Data Science** – Supervised and unsupervised learning algorithms, regression, classification, clustering – *k*-Nearest Neighbors, Support Vector Machines, Logistic Regression, Decision Trees, Random Forest, Naïve Bayes, Stochastic Gradient Descent **Tutorial 6:** Image Processing with Python Due: Homework 3

Week 7: Oct. 6	**Model Engineering Pipelines** **Lecture 13: Ensemble Models** – Voting, bagging, boosting, stacking ensembles **Lecture 14: Artificial Neural Networks with Keras-TensorFlow** – Activation functions, losses, backpropagation, fully-connected layers, classification and regression **Tutorial 7:** TensorFlow **Quiz 2**
Week 8: Oct. 13	**Quiz 3**
Week 9: Oct. 20	**Model Engineering Pipelines** **Lecture 15: Convolutional Neural Networks with Keras-TensorFlow** – Convolutional layers, models for image classification, transfer learning, fine-tuning pretrained models **Lecture 16: Model selection, Hyperparameter Tuning, Callbacks** – Grid search, cross-validation, model evaluation, callbacks for model monitoring **Tutorial 8**: PyTorch
Week 10: Oct. 27	**Model Engineering Pipelines** **Lecture 17: Artificial Neural Networks with PyTorch** – Data loaders, model definition, training, and evaluation, custom datasets, model saving and loading **Lecture 18: Natural Language Processing** – Preprocessing text data, tokenization, representation of word groups, sequence model approach, word embeddings **Tutorial 9:** Bash Scripting
Week 11: Nov. 3	**Model Engineering Pipelines** **Lecture 19: Transformer Networks** – Self-attention mechanism, multi-head attention, positional encoding, encoder and decoder blocks, fine-tuning pretrained models **Lecture 20: NLP with Hugging Face** – Pipelines for NLP tasks, tokenizers, datasets, models Due: Homework 4
Week 12: Nov. 10	**Model Engineering Pipelines** **Lecture 21: Large Language Models** – Fine-tunning a pretrained model, low-rank adaptation (LoRA), prefix tuning, prompt tuning – Mixture of expert models, retrieval-augmented generation (RAG), multi-modal models **Tutorial 10:** Git and Version Control **Quiz 4**

Week 13: Nov. 17	**Model Engineering Pipelines** **Lecture 22: Reasoning Large Language Models** – Test-time compute scaling, chain-of-thought, search against verifier models, reinforcement learning-based reasoning **Lecture 23: Agentic AI** – Planning, tool use, memory, reflection, agent frameworks and architectures, applications **Tutorial 11:** TensorFlow Due: Homework 5
Week 14: Dec. 1	**Deployment Pipelines** **Lecture 24: Introduction to Data Science Operations (DSOps)** – DS project life cycle, levels of automation, model deployment, model serving in production **Lecture 25: Deploying Projects as Web Applications** – Intro to Flask, FastAPI, Django, creating RESTful API to serve data and model predictions, models for tabular data and image classification Serving **Quiz 5**
Week 15: Dec. 8	**Deployment Pipelines** **Lecture 26: Deploying Projects to the Cloud** – Intro to Azure Machine Learning, no-code ML, code-based ML, training deep learning models with Azure ML **Lecture 27: Reproducible Projects, Docker Containers, Kubernetes** – Intro to Docker and containerization, building and managing Docker containers, Kubernetes for container orchestration and management Due: Homework 6
Week 16: Dec. 15	**Quiz 6**

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, and 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 concerns. 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), 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 the 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.

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 the integrity of tribal voices.