Course Syllabus – CBB 510

Course Information

Course Number: CBB 510 SP24
Course Name: Advanced Data Science
Term: SP 2024
Start Date: 04/02/2024
End Date: 06/21/2024
Credits: 3.0

Meeting Days / Times

Tuesdays and Thursdays, 1:45-3:15pm PT / 4:45-6:15pm ET
(See Calendar in Canvas for the most up-to-date schedule.)

Location

CA: Graduate Office Dining Room (Hazen Theory Building)
FL: B214

Course Managers

Role	Last Name	First Name	Email Address
Course Director	Quer	Giorgio	gquer@scripps.edu
Course Director	Wu	Chunlei	cwu@scripps.edu
TA	Bruciaferri	Niccolo	nbruciaferri@scripps.edu
TA	Gonzalez Cavazos	Carolina	agonzalez@scripps.edu
TA	Nagaraja	Shashank	snagaraja@scripps.edu

Course Description

This course will focus on the recent advances in data science and their application in biomedical research, in which the bioinformatics and data science approaches have played an essential role. The course is designed as a series of lectures by either course directors or invited speakers, each of which will focus on a particular topic per week and followed by a journal-club style discussion or a hand-on coding exercise on the same topic. The covered Data Science topics are generally organized in two dimensions: 1) Data Science techniques, covering topics such as Machine-learning algorithms, NLP/LLM, Knowledge graphs, Alphafold, as well as general Python programming skills; 2) Biomedical Data Science scenarios, covering topics such as

scRNA-Seq, Proteomics data. The primary goal of this course is to enhance students' knowledge foundation and practical skills by exposing them to a broad diversity of state-of-art Data Science topics.

The target students are those who have finished introductory bioinformatics training (e.g., finished Applied Bioinformatics or Introduction to Data Science or Introduction to Biostatistics course), and look for the continued advancement of their computational skills and knowledge. This course also prepares students for applying data science approaches in their own research area, as well as writing and presenting research projects in the computational field. Coding exercises will be done in Python with possibly few exceptions (e.g. might use R in some invited lectures), so basic Python programming skills are required and will be summarized at the beginning of the class.

Program Learning Outcomes

By the end of the program, students will have accomplished these objectives:
PLO1: Original Research – graduate students are expected to develop the skills critical for generating high-quality research output. This would include absorbing, recalling, and contextualizing scientific knowledge, evaluating scientific information and data, creating testable hypotheses and investigating hypotheses, mastering scientific tools and techniques, displaying ethical behavior, and receiving and giving feedback.
PLO2: Communication – graduate students are expected to demonstrate the oral, written, and media skills to effectively communicate the impact of a study or a body of work to the greater scientific community and to the public at large using a number of methods.
PLO3: Critical Thinking – graduate students are expected to develop a self-directed process to analyze information, form opinions or judgments, and use this process to improve the quality of their scientific thoughts, navigate problems, and make informed decisions.
PLO4: Intellectual Curiosity – graduate students are expected to acquire the capacity to build their intellectual curiosity and demonstrate problem solving approaches that serve their professional growth and ability to impact a field.
PLO5: Career and Professional Development – graduate students are expected to develop a variety of transferable skillsets throughout their graduate experience, including management and leadership, inclusiveness, resilience, scientific rigor, collaboration, accountability, time management, teamwork, networking, and career planning.

For a detailed description of each outcome and specific success indicators, please refer to this web page: https://education.scripps.edu/graduate/doctoral-program/.

Course Learning Outcomes

Upon completion of this course students will be able to:
CLO1: Understanding and writing basic Python code, managing datasets and manipulating text and image data.

CLO2: Basic understanding of major AI tools in healthcare, including deeper understanding of Large Language Models and Alphafold.
CLO3: Understanding the basics of the biomedical data modeling and knowledge integration.
CLO4: Obtaining basic knowledge covering diversified techniques and use cases in the field of data science in life sciences.

Background Preparation (Prerequisites)

Completion of two "Applied Bioinformatics" courses, STBIO 400 and STBIO 440, are required, or permission by instructor(s) (for example, justified from the relevant research experience).

Course Materials

The material (slides of the lectures, papers to be reviewed during journal club, Jupyter notebooks and other AI tools) will be provided to the student during lectures. There is not a suggested textbook.

Class Format, Expectations and Logistics

The class will meet twice a week for 1.5 hours each class.
Every week there will be one traditional lecture and one applied class. The traditional lecture will be lead by the Instructor or an Invited Speaker on a specific topic on AI or data science, leaving room for questions during the class and at the end of the class.
The applied class will be of different format. The first week will be dedicated to the general Python programming skills, so that everyone is prepared for the future coding exercises. When a coding exercise is planned, a set of example code will be discussed and then followed with a session of coding practices. When a journal club is planned, at least two research or review papers in the focused topic will be assigned at the beginning of the week. Students will be separated into two pre-assigned groups and each group will have a lead presenter. Each group presenter will have 30-40 minutes to present their assigned paper and answer the questions from the instructor and the whole class. To guide the discussion, the instructor may provide a set of questions along with the paper assignments. Students are responsible for reading these materials before the discussion. Each student is also required to put down questions they want to discuss at the class.

There will be one final coding assignment at the end of the course. The final assignment will be released at least one week before the final week, and it's due by the end of the last week.

Attendance Statement

Students are expected to attend all classes. Students who are unable to attend class must seek permission for an excused absence from the teaching assistant (via direct message on Slack to Shashank). Unapproved absences or late attendance for three or more classes may result in a lower grade or an "incomplete" for the course. If a student has to miss a class, they should

arrange to get notes from a fellow student and is strongly encouraged to meet with the teaching assistant to obtain the missed material.

Scientific and Professional Ethics

The work you do in this course must be your own. Feel free to build on, react to, criticize, and analyze the ideas of others but, when you do, make it known whose ideas you are working with. You must explicitly acknowledge when your work builds on someone else's ideas, including ideas of classmates, professors, and authors you read. If you ever have questions about drawing the line between others' work and your own, ask the course professor who will give you clear guidance. Exams must be completed independently. Any collaboration on answers to exams, unless expressly permitted, may result in an automatic failing grade and possible expulsion from the Graduate Program.

Technology Requirements and Support

For issues related to Canvas, please contact the Graduate Office by email at: gradprgm@scripps.edu or by phone at: 858-784-8469.

Course Grading

Grading is in accordance with the academic policies of the Skaggs Graduate School. The breakdown of grading is as follows:

- Coding assignments: 25%
- Journal club assignments: 25%
- Final assignment: 30%
- Participation: 20%

Letter Grade	Percent	GPA	Description
A	93-100	4.00	Outstanding achievement. Student performance demonstrates full command of the course subject matter and evinces a high level of originality and/or creativity that far surpasses course expectations.
A-	90-92	3.67	Excellent achievement. Student performance demonstrates thorough knowledge of the course subject matter and exceeds course expectations by completing all requirements in a superior manner.
B+	87-89	3.33	Very good work. Student performance demonstrates above-average comprehension of the course subject matter and exceeds

			course expectations on all tasks as defined in the course syllabus. There is notable insight and originality.
B	83-86	3.00	Satisfactory work. Student performance meets designated course expectations and demonstrates understanding of the course subject matter at an acceptable level.
B-	80-82	2.67	Marginal work. Student performance demonstrates incomplete understanding of course subject matter. There is limited perception and originality.
C+	77-79	2.33	Unsatisfactory work. Student performance demonstrates incomplete and inadequate understanding of course subject matter. There is severely limited or no perception or originality. Course will not count toward degree.
C	73-76	2.00	Unsatisfactory work. Student performance demonstrates incomplete and inadequate understanding of course subject matter. There is severely limited or no perception or originality. Course will not count toward degree.
P	73-100	0.00	Satisfactory work. Student performance demonstrated complete and adequate understanding of course subject matter. Course will count toward degree.
F	0-72	0.00	Unacceptable work/Failure. Student performance is unacceptably low level of knowledge and understanding of course subject matter. Course will not count toward degree. Student may continue in program only with permission of the Dean.
I		0.00	Incomplete is assigned when work is of passing quality but is incomplete for a pre-approved reason. Once an incomplete grade is assigned, it remains on student's permanent record until a grade is awarded.
W		0.00	Withdrew from the course with Dean's permission beyond the second week of the term.

- All courses will be recorded and maintained in the student's permanent academic record; only courses that apply towards the degree will appear on the academic transcript. Non-credit or audited courses will not appear on the transcript.
- 4 core courses taken for a letter grade (pass = B- or higher for a core course)
- 2 elective courses taken pass/fail (pass = A, B, C for an elective)

Because students are encouraged to take electives outside their area of expertise, a "C" letter grade is passing.

Course Summary:

Date	Details
Tue Apr 2, 2024	Introduction (Quer)
Thu Apr 4, 2024	Python setup, python, exercises (Quer)
Tue Apr 9, 2024	Large Language Model (LLM): introduction and its use with Azure (Carletti)
Thu Apr 11, 2024	Journal club + Assignment on LLM
Tue Apr 16, 2024	Single-cell sequencing analysis (Doug Evans)
Thu Apr 18, 2024	Journal club
Tue Apr 23, 2024	Python programming - standard packages for data science (Quer)
Thu Apr 25, 2024	Coding practices (with homework)
Tue Apr 30, 2024	Biomedical DB and APIs (Wu)
Thu May 2, 2024	Coding Practices
Tue May 7, 2024	Unimodal, multimodal AI and LLM in cardiovascular care (Quer)
Thu May 9, 2024	Journal Club
Tue May 14, 2024	Proteomics data analysis (Yates)
Thu May 16, 2024	Journal Club
Fri May 17, 2024	Commencement
Tue May 21, 2024	Text-mining and NLP: from text analysis in python to LLMs (Quer)
Thu May 23, 2024	Coding practices (with homework)
Mon May 27, 2024	No Class (Memorial Day)
Tue May 28, 2024	Knowledge Graph and Ontology (Wu)
Thu May 30, 2024	Journal Club
Tue Jun 4, 2024	Prediction of protein structure with Alpha-fold (Forli)
Thu Jun 6, 2024	Journal Club
Tue Jun 11, 2024	Python programming, or other topic (Wu)
Thu Jun 13, 2024	Coding Practices
Tue Jun 18, 2024	Final in-class coding assignment
Wed Jun 19, 2024	No Class (Juneteenth)
Thu Jun 20, 2024	TBD