**DATA SCIENCE: Models, Processes and Applications**

Computer Science 4150/5150 (offered initially as CS 4900/5900)

**Course description**:

Students who complete this course will gain a thorough understanding of computational techniques and processes for data science by implementing them and by applying them to solve real world problems. Students will gain not only an understanding of data science models and methods, but they also will learn how to perform analyses to answer domain research questions and to effectively communicate insights that result from their analyses.

**Learning Objectives[[1]](#footnote-1)**:

Students will be able to clean, preprocess, and explore real world data sets.

Students will be able to build computational models (perform computational modeling) for graph analytics and for classification.

Students will be able to characterize the statistical significance of their computational models.

Students will be able to visualize and interpret the significance of the computational models in authentic domains.

**Assessment[[2]](#footnote-2)**:

20% - Quizzes

20% - In class activities

25% - Midterm project

35% - Final project

# Quizzes: blackboard will be based on assigned readings and videos from online sources.

# Grading[[3]](#footnote-3)

Makeup activities and project deadline extensions will not be granted, except for legitimate reasons (as defined in the student handbook).

Participation in all class sessions is mandatory.

Grades will be assigned as follows:

A-, A 90-93, 94-100%

B-, B, B+ 80-83, 84-86, 87-89%

C-, C, C+ 70-73, 74-76, 77-79%

D-, D, D+ 60-63, 64-66, 67-69%

**Instructor**: Dr. L. R. Welch, Stocker 354, [welch@ohio.edu](mailto:welch@ohio.edu). Tuesday & Thursday 10:00am.

|  |  |
| --- | --- |
| **Week** | **Topics** |
| **1** | intro; 3D genomics; data cleaning, preprocessing, and exploration |
| **2** | 3D interactomics; pattern inference |
| **3** | distance and similarity metrics |
| **4** | clustering |
| **5** | midterm project demo - phase I |
| **6** | feature identification, cell cycle |
| **7** | feature selection, state estimation |
| **8** | midterm project demo - phase II |
| **SPRING**  **BREAK** |  |
| **9** | co-segregation, normalization, visualization, exploratory statistics |
| **10** | final project demo - phase I |
| **11** | network centrality |
| **12** | network community discovery, visualization |
| **13** | final project demo - phase II |
| **14** | feature mapping and selection,  model interpretation |
| **15** | final project |
| **FINAL EXAM WEEK** | final project demo - phase III |

1. Implements the following ABET outcomes: (b) An ability to analyze a problem, and identify and define the computing requirements appropriate to its solution. (c) An ability to design, implement, and evaluate a computer-based system, process, component, or program to meet desired needs. [↑](#footnote-ref-1)
2. Accommodations for Students with Disabilities: Any student who suspects s/he may need an accommodation based on the impact of a disability should contact the class instructor privately to discuss the student’s specific needs and provide written documentation from the Office of Student Accessibility Services. If the student is not yet registered as a student with a disability, s/he should contact the Office of Student Accessibility Services. [↑](#footnote-ref-2)
3. Individuals performing plagiarism, copying (e.g., using software code written by another student) and other forms of *academic misconduct* (see student handbook for further details) will receive an ‘F’ in the course and referral to Ohio University Judiciaries. [↑](#footnote-ref-3)