Course Syllabus

Course Information

Title: Data Science

Class number: CS 479/579

Class time: Tue./Thur. 2:00PM – 03:15PM PT, Spring of 2024

Class location: MCCL #209 @ Moscow; IFCHE #303 @ Idaho Falls; BLDB #211 @ CdA

Zoom link: **https://uidaho.zoom.us/j/83101111099 (https://uidaho.zoom.us/j/83101111099)**

Instructor: Xiaogang Ma, Associate Professor at CS Dept.

Instructor contact: Office: JEB#230; Email: max@uidaho.edu; Office phone: 208.885.1547

Office hours: Tue./Thur. 4:00PM-5:00PM PT; In Person, or Zoom: **https://uidaho.zoom.us/j/2018585704 (https://uidaho.zoom.us/j/2018585704)**

Synopsis

Data science is advancing the conduct of science in individual and collaborative works. Data science combines aspects of data management, library science, computer science, and physical science using supporting cyberinfrastructure and information technology. Key methodologies in application areas based on real research experience are taught to build a skill-set that enables students to handle each stage in a data lifecycle, from data collection, analysis, archiving, to data discovery, access and reuse.

Course Outline and Description

Prerequisites:

- CS 360 Database Systems

Learning Objectives:

1. **Understand** the fundamental concepts in data science, such as data, information, knowledge, metadata, data lifecycle, data management and data analysis, as well as the inter-relationships

among those concepts. Through independent learning and collaborative discussion with classmates from different disciplines, student will attain an integrative overview of data science.
2. **Develop and Demonstrate** skills for steps in a data lifecycle, including data collection, management, analysis and product generation. Students will learn and improve their skills with real world examples and work in groups of classmates from multi-disciplinary backgrounds.
3. Be proficient in the **publication and communication** of data and information products. Students will learn state-of-the-art technologies in data visualization and work in groups to prepare and present the outputs of course projects.
4. Understand the **diversity and ethics** in the conduction of data science. Students will discuss and learn ethical concepts such as neutral perspective, privacy, intellectual property, accountability and responsibility in data science. They will also apply those ethical guidelines in their course projects.
5. Apply principles of **respect** in collaboration and interaction with others in course study, group discussion, and course projects.

Course Schedule:

1. Basic concepts in data science
2. Data collection, stewardship and preservation

- Assignment 1: Data collection proposal, **10%** of grade

3. Data formats and standards
4. Class exercise - data collection (individual)

- Assignment 2: Data presentation, **15%** of grade

5. Open Data and the Web of Data
6. Class presentation - present your data (individual)

- Assignment 3: Data cleansing and EDA, **10%** of grade

7. Data science process and exploratory data analysis
8. Algorithms and class exercise - group course project definitions

- Assignment 4: Group work on a data science project
  - Progress presentation, **5%** of grade
  - Final presentation, **10%** of grade
  - Written report, **30%** of grade

9. Interdisciplinary data science: Examples

[Spring Recess Week]

10. Data mining
11. Data quality, uncertainty, and bias
12. Class presentation - group course project progress report
13. Data visualization

14. Data workflow
15. The data ecosystem
16. Final group project presentations

Assessment Method

Learning Activities:

- Attendance and participation in lectures
- Reading and writing assignments
- Class exercises and group discussion/presentation
- Course project

Grading Criteria:

The grading (A/B/C/D/F) is based on the student learning activities.

- 20% Attendance
- 35% Class exercise and assignment
- 45% Class project report and presentation