

LEADING Boot Camp: WEEK 2

- Please scroll down to the weekly folders.
- Please also note that the OpenRefine asynchronous session accessible in the Week 2 folder.

Monday, June 12, 2023

10:30-12:15 PM Supervised Learning (Professor Lei Wang)

Overview of Machine Learning

Frequently utilized supervised learning methods

Evaluation of the supervised learning

12:15-1:15 PM: BREAK

1:15-2:15 PM Command Line and Scripting in the Data Science Pipeline (Professor Weimao Ke)

Shell and command lines (for Linux, Mac, and Windows)

Piping and redirection, power to combine tools

Case study and demo

2:15-2:30 PM: BREAK

2:30-3:30 PM EDT Data Preservation /Web Archiving (Professor Mat Kelly)

Archival crawling

Access and replay

Formats and metadata

Wednesday, June 14

10:30-11:30 AM Big Metadata and Data Curation (Professors Jane Greenberg and Alex Poole)

Metadata, big metadata, ontologies

Linked open data

Data lifecycle and curation, principles

11:30 AM-12:15 PM Icebreaker Session: Get to know your cohort (Jane Greenberg and Alex Poole)

12:15-1:15 PM EDT: BREAK

1:15-3:30 PM Data Management: SQL and NoSQL (Professor Il-Yeol Song)

Understand SQL as a database language

Use SQL for data science

Understand differences between SQL and NoSQL

◆◆◆ Data Cleaning and Transformation with Open Refine (Jon Wheeler, Data Curation Librarian, University of New Mexico)

Additional
Asynchronous
Session

- Inspecting data through faceting, filtering, and clustering
- Data cleaning and transformation
- Scripted workflows and automation