Sungkyunkwan University (SKKU) International Summer Semester (ISS) 2025

Health Data Analytics & AI

Prof. Javed Mostafa, University of Toronto

The

SHORT COURSE DESCRIPTION

What diseases tend to occur more frequently in neighborhoods near chemical plants? How can an oncologist view and understand the long-term progression of cancer in one of her patients? The health analytics course will introduce students to a broad range of informatics and data analytics methods used to answer these kinds of questions. The course will focus on building students' practical skills for processing and analyzing both structured and unstructured clinical data by taking them through the implementation of **specific statistical methods and basic machine learning methods** (applied AI techniques) to understand trends in individual and population health. By taking this course, students will become aware of the range of tools in a data analyst's and applied AI expert's toolbox and learn which ones to use to answer their own clinical or population health research questions.

READING MATERIALS

Most materials for the course will be available on the course Sakai site and the interactive textbook in the GoogleCollab site. Information for the required textbook is listed below.

- Textbook: Berman, J. J. (2018). Methods in Medical Informatics: Fundamentals of Healthcare Programming in Perl, Python, and Ruby. CRC Press.
- Selected interactive chapters used in the course are hosted on a cloud platform and we plan to share the URL with you on May 17th during the first interactive class session (we call this site: GoogleCollab). You'll be given instructions for accessing a password-protected read-write version to use for the course during the first meeting on the first class day.

COURSE REQUIREMENTS AND GRADING

Total grade score achievable is 100. The course grade is determined via the following weighting:

Forum posts – 35%
Project Plan – 5% Group project:
Assignment 1 – 10%
Assignment 2 – 15%
Assignment 3 – 15%
Project Completion & Presentation – 20%

Participation grades will account for forum posting and synchronous session engagement. Responding to other students' questions in the forum will also positively impact your participation grade. Additional information on the project development and deliverables will be shared in the class.

A final grade score 60 or above will be considered a Pass.

COURSE SCHEDULE

– WEEK I –

Monday (30 June)

Introduction to health analytics and informatics. Review of syllabus and class requirements.

Tuesday (1 July)

Utility Scripts: Complete Chapter 2 on GoogleCollab

Wednesday (2 July)

Utility Scripts: Complete forum post and exercises

Thursday (3 July)

Viewing and Modifying Images: Complete Chapter 3 on GoogleCollab

– WEEK II –

Monday (7 July)

Viewing and Modifying Images: Complete forum post and exercises

Tuesday (8 July)

Indexing & MeSH: Complete Chapter 4 GoogleCollab

Wednesday (9 July)

Indexing & MeSH: Complete Chapter 5 on GoogleCollab

Thursday (10 July)

Controlled Vocabularies (ICD) & NLM: Complete Chapter 6 on GoogleCollab

– WEEK III –

Monday (14 July)

Controlled Vocabularies (ICD) & NLM: Complete Chapter 9 on GoogleCollab

Tuesday (15 July)

Working with External Data: API access to and Local Data extracted from Pubmed: Chapter 9

Wednesday (16 July)

Leveraging Taxonomies and Classification Schemes: Chapter 10

Thursday (17 July)

Group Project Progress and Mock Demos

– WEEK IV–

Monday (21 July)

Structuring Data and Manipulating Structured Data: Complete Chapter 18 GoogleCollab

Tuesday (22 July)

Mining Information: Complete Chapter 14 GoogleCollab

Wednesday (23 July)

Project Presentations