Data Science in Drug Discovery, Health and Translational Medicine
Course Number(s): INFO I-590
Instructor: Joanne S. Luciano
January 2017

Instructor Information:
Name: Joanne Luciano
Title: Visiting Associate Professor
Office Address: 711 N Park Ave
Email: jluciano@indiana.edu
Webpage: N/A
Office Number: 104
Social Media Handles: Twitter: joanneluciano
Office Hours: TBD

Lead AI(s) Information:
Name: Jeremy J Yang
Title: Research Scientist (UNM), PhD student (IU)
Email: jejyang@indiana.edu
Webpage: http://homes.soic.indiana.edu/jejyang/
Office Hours: TBD

Course Description/Overview:
With exploding healthcare costs, greater longevity and widespread health challenges of diabetes, obesity, cancer, and cardiovascular disease, medicine and healthcare will be a primary scientific and economic focus for the next two to three decades. Informatics and data science offer the promise of a level of understanding of health, disease and treatment on a scale never before imagined. This course will provide an overview of the healthcare ecosystem and the data science techniques that are being used in the drug discovery, healthcare and translational medicine domains. The course starts with an overview of data science, the healthcare ecosystem, and the data and metadata in this domain. With this foundation, the remainder of the course is organized around three questions: how can data science help researchers find new drugs and reuse old ones? How can data science help doctors and other clinicians treat patients better? And how can data science help us all lead healthier lives?
The course is broken down into modules, based on the foundational required knowledge and these three questions. Each week of the course will focus on 1-3 topics. Each module will have four parts: Video(s), which gives an overview of the topic(s); Learning Goals that list what you should aim to know after completing the module; Learning Tasks that all students should complete in addition to watching the video(s), and Going Deeper that gives resources for advanced students and those that want to go deeper into the material.

Course Goals:
- The students will learn about the theoretical framework of data science and its tools.
- The students will learn about the healthcare ecosystem, and the roles data science plays in drug discovery, healthcare and translational medicine.
- The students will learn about biomedical data, metadata, and challenges unique to health data.
- The students will learn effective strategies for integrating, analyzing and presenting data.
- The students will develop skills in using Python, R, and Tableau.
- The students will learn how to use questions to guide themselves through data acquisition, integration, analysis, and presentation using data visualization and story.
- The students will learn the historical and current paradigms of the drug discovery process
- The students will learn the relevance and value of bioinformatics to medicine, drug discovery and healthcare.
- The students will learn the relevance and value of cheminformatics to medicine, drug discovery and healthcare.

Course Learning Objectives:
At the end of this course students will be able to:
- Articulate the major components of the healthcare ecosystem and the common tools and analyses used in each.
- Demonstrate understanding of the most important metadata standards, where and how they are used to facilitate health analytics.
- Integrate heterogeneous biomedical and healthcare data using biomedical metadata standards.
- Produce a report with evaluations and visualizations to support your conclusion based on public data downloaded from the Web.
- Integrate multiple sources, clean, analyze, and present results using R, Python and Tableau.
- Evaluate the quality of healthcare data and metadata for use in analysis.
- Build a prototype health app for a mobile device.
- Critically analyze drug discovery processes in relation to historical and current paradigms, especially relative to informatics and data science.
- Develop data science skills and knowledge in utilizing online bioinformatics data resources and analysis tools.
- Develop data science skills and knowledge in utilizing online cheminformatics data resources and analysis tools.
Course Prerequisites:
Students should have had an introductory course in programming or equivalent. Familiarity with the Python language, the R statistical package, basic machine learning algorithms is desirable. Background in a healthcare field is helpful but not required.

Course Books/Readings/Tutorials/Other Resources:
Books:
R for Medicine and Biology; Paul D. Lewis, PhD, Swansea University

Python for Bioinformatics; Jason Kinser, DSc, George Mason University
http://www.jblearning.com/catalog/9780763751869/

Course Communication Channel(s):
Canvas Discussion Forum; Email listserv; Zoom for office hours.

Course Policies:
- This course follows Indiana University's academic calendar for Spring 2017.
- Students are expected to carry out all activities within each module. Each module has some small number of activities associated with it.
- Students are expected to maintain more than one copy of their assignments in case of equipment failures.
- It is expected that a student will put in 5-7 hours a week every week into the course which includes time spent in readings, reflections, and engaging with instructional content.

Incomplete/Drop Policies:
Incomplete Policy:
Circumstances Permitting Incompletes:
(Approved: Faculty Council 11/5/52, amended 2/19/63)
The grade of Incomplete used on the final grade reports indicates that the work is satisfactory as of the end of the semester but has not been completed. The grade of Incomplete may be given only when the completed portion of a student's work in the course is of passing quality. Instructors may award the grade of Incomplete upon a showing of such hardship to a student as would render it unjust to hold the student to the time limits previously fixed for the completion of his/her work.

Departmental Records:
Each academic unit shall maintain a record of Incomplete grades recorded in its courses. This record, completed by the instructor, should include (l) the name of the student and the student's
identification number, (2) the course number, section number, and hours of credit, (3) semester and year of enrollment, (4) the signature of the instructor, (5) a brief statement of the reason for recording the Incomplete, and (6) an adequate guide for removal of the Incomplete grade (with a suggested final grade) in the event of the departure or extended absence of the instructor from the campus.

Removal of Incompletes:

Methods. A grade of Incomplete may be removed (a) by the student completing the course within the time limit and the instructor sending the appropriate Removal- of-Incomplete form to the Office of the Registrar, and (b) by the dean of the student's school authorizing the change of Incomplete to W.

Limits. The time allowed for the removal of an Incomplete is one calendar year from the date of its recording, except that the dean of the student's college or school may authorize adjustment of this period in exceptional circumstances. By assigning an Incomplete an instructor implicitly authorizes and requires the "I" to be changed to an "F" at the end of the appropriate time period, if that instructor does not otherwise act to remove the "I". The Registrar will automatically change the "I" to "F" at the end of the appropriate time period except when an adjustment of the period has been authorized or the student has received a degree since that date. Both the student and the instructor in whose course the student received the Incomplete will be notified of this change of grade.

A student may not re-enroll in a course in which a grade of Incomplete has been recorded. The student may be denied the right to make up an Incomplete if it seems to the unit dean and the instructor that it is impractical for the student to complete the course. In this case, the student should be given the opportunity to withdraw from the course.

Absence from Final Examinations:

(Approved: Faculty Council 11/5/52)

Where the grade of Incomplete is given because the student missed the final examination, he or she shall be allowed to remove the Incomplete by taking the examination only if he or she has followed the regular procedure to have his/her absence excused and the Committee on Absence has notified the instructor that the student may be permitted to take the examination. If the Committee on Absence, under the Dean for Student Services, determines that the reason for the student's absence is not satisfactory, it should inform the instructor that the grade of Incomplete should be changed to a grade of "F". Where the Incomplete was received because of absence from the final examination, students may prefer to receive a grade of W instead of taking the examination. In such cases, the Dean will not approve the grade of W unless the Committee on Absence has approved the reason for absence.

Drop Policy:

Before you drop or add a class, make sure you understand the rules. Your timing makes all the difference when it comes to whether or not you're hit with extra fees or a bad grade.
Dropping or adding classes can affect your financial aid, scholarship status, or tuition cost. If you do decide to drop or add a class, be sure to talk with your academic advisor first. If you're worried about what will happen with your finances, contact us. We'll talk you through it.

**Making changes during the continuous drop/add period:**

**Continuous drop/add** is available from the beginning of early registration through the Sunday following the first week of classes. If you drop a class during this time, it won’t show up on your transcript or grade reports. So if you want to avoid a W (“Withdrawn”) grade, make sure to drop by this **deadline**.

**Schedule adjustment fees:**

You'll be charged a drop/add access fee of $8.50 each day you make a schedule change after two business days from the date you register for the term.

If you drop courses after the first week of classes, you'll be charged an additional fee for each course you drop.

**Student Integrity Policy:**

Students are expected to conduct themselves in a manner befitting their status as a student of a respected and distinguished institution of higher education. In college courses, we are continually engaged with other people’s ideas: we read them in texts, hear them in lecture, discuss them in class, and incorporate them into our own writing. As a result, it is very important that we give credit where it is due. Plagiarism is using others’ ideas and words without clearly acknowledging the source of that information.

See [http://www.indiana.edu/~wts/pamphlets/plagiarism.shtml](http://www.indiana.edu/~wts/pamphlets/plagiarism.shtml) for help in addressing plagiarism in your own work.

Articles and data posted to Canvas for this course are for the convenience of students only and should not be shared outside the course.

**Students with Disabilities Accommodations:**

**Arranging Accommodations:**

Once you have submitted a request for service, provided appropriate documentation verify your disability, and met with your DSS Coordinator, it is time to arrange your reasonable accommodations with your instructors.

**What is a Reasonable Accommodation?**

A reasonable accommodation is any modification or adjustment that will enable a qualified student with a disability to participate in a course, program, activity or service. Reasonable accommodations assure that individuals with a disability have rights and privileges equal to students without disabilities. Reasonable accommodations may include academic adjustments, auxiliary aids or adaptive technology, services, or modifications for facilities.

**Academic Accommodation Memorandum (Memo):**

The DSS office prepares an Academic Accommodation Memorandum (Memo) for each student receiving accommodations. This document explains the reasonable accommodations to be provided to you as a student of IU to your instructors. It is your responsibility to give a copy of our memo to each of your instructors at the beginning of each academic term.
Request an Academic Accommodation Memo.

Submitting Memos to Your Instructors:
As a student receiving accommodations, you are required to meet with each of your professors in person either during office hours or by individual appointment. Electronic web meetings may be substituted for online students. It’s necessary for you to meet your professor early in the semester - within the first two weeks if possible. By the time you leave your meetings with professors, you should have an understanding of how the professor will handle your accommodations during the semester.
If you have any questions about the memo or any part of the DSS system and process, don’t hesitate to reach out to your coordinator.

Course Roadmap:
Section 1. Introduction to Data Science and the Healthcare Ecosystem

1. Data Science
   a. Concepts
   b. Workflow
   c. Questions & Tools
   d. Evaluation and Reporting
2. Healthcare
   a. Ecosystem
   b. Data
   c. Metadata
3. Health Web Science
   a. The role of the web in healthcare
   b. 21st century medicine

Section 2: Q1 How can data science help researchers find new drugs and reuse old ones?

1. The drug discovery process
   a. Epochs and paradigms of drug discovery
   b. Integrative data science, a new paradigm
2. Bioinformatics
   a. What is bioinformatics? Genomics and proteomics.
   b. The Central Dogma of Molecular Biology
   c. The Human Genome Project
   d. Genomics and precision medicine
3. Cheminformatics
   a. What is cheminformatics?
   b. Chemicals represented as molecular graphs
   c. Chemical databases
   d. Exploring chemical space for new drugs

Section 3: Q2 How can data science help providers treat patients better?

1. Providers / Hospitals (institutions) EHR, Clinical Trials, Precision Medicine
2. Apps for Doctors, Nurses, Clinicians

Section 4: Q3 How can data science help us all lead healthier lives?

1. Patient Focused. (Quantified Self, electronic diaries, App development, Your Medical Mind)
2. Crowdsourcing, online communities, ePatients, patient advocates
3. Personal Genomics

Topic Outline:

<table>
<thead>
<tr>
<th>Week #</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Module</strong>: Course Overview, Data Science Concepts</td>
</tr>
<tr>
<td>2</td>
<td><strong>Module</strong>: Healthcare Part 1: Ecosystem, Data</td>
</tr>
<tr>
<td>3</td>
<td><strong>Module</strong>: Healthcare Part 2: Metadata</td>
</tr>
<tr>
<td>4</td>
<td><strong>Module</strong>: Health Web Science: The role of the web in healthcare and 21st Century Medicine</td>
</tr>
<tr>
<td>5</td>
<td><strong>Module</strong>: Q1-1 The Drug Discovery Process: History, epochs and paradigms of drug discovery</td>
</tr>
<tr>
<td>6</td>
<td><strong>Module</strong>: Q1-2 Bioinformatics: genomics, proteomics, the Central Dogma, the Human Genome Project and its offspring</td>
</tr>
<tr>
<td>7</td>
<td><strong>Module</strong>: Q1-3 Bioinformatics: medicine as a molecular information science</td>
</tr>
<tr>
<td>8</td>
<td><strong>Module</strong>: Q1-4 Cheminformatics: molecular graphs and databases, the vastness of chemical space, the search for new drugs, a high dimensionality challenge</td>
</tr>
<tr>
<td>9</td>
<td><strong>Module</strong>: Q2-1 Providers / Hospitals (institutions) EHR, Clinical Trials, Precision Medicine</td>
</tr>
<tr>
<td>10</td>
<td><strong>Module</strong>: Q2-2 Apps for Doctors, Nurses, Clinicians</td>
</tr>
<tr>
<td>11</td>
<td><strong>Module</strong>: Q2-3 Government (Policy, Public Health (HealthMap, Google Flu), role), Privacy, Industry, Insurance Policy, (smoking, FDA)</td>
</tr>
<tr>
<td>12</td>
<td><strong>Module</strong>: Q3-1 Patient Focused. (Quantified Self, electronic diaries, App development, Your Medical Mind)</td>
</tr>
<tr>
<td>13</td>
<td><strong>Module</strong>: Q3-2 Crowdsourcing, online communities, ePatients, patient advocates</td>
</tr>
</tbody>
</table>
### Possible Assignments/Project:

**Example 1:** Take a public dataset health dataset (perhaps smoking data) and create a graphic of it. Maybe someone looking for an environmentally safe place to live.

**Example 2:** Using online resources such as Kyoto Encyclopedia for Genes and Genomes (KEGG), and Online Mendelian Inheritance in Man (OMIM), and PubChem, find genes associated with a disease, corresponding proteins, and bioassays, and bioactive drug leads.

Other projects and assignments TBD.
Background: Data Science and the Healthcare Ecosystem (Background)

- Week 1 Module 1: Course Overview, Data Science Concepts
- Week 2 Module 2: Healthcare Part 1: Ecosystem, Data
- Week 3 Module 3: Healthcare Part 2: Metadata, Ontologies
- Week 4 Module 4: Health Web Science: The role of the Web in healthcare and 21st Century Medicine

Science Applications in Drug Discovery
Q1: How can data science help researchers find new drugs and reuse old ones?

- Week 5 Module 5: Q1-1 The Drug Discovery Process: History, epochs and paradigms of drug discovery
- Week 6 Module 6: Q1-2 Bioinformatics: genomics, proteomics, the Central Dogma, the Human Genome Project and its offspring
- Week 7 Module 7: Q1-3 Bioinformatics: medicine as a molecular information science
- Week 8 Module 8: Q1-4 Cheminformatics: molecular graphs and databases, the vastness of chemical space, the search for new drugs, a high dimensionality challenge

Data Science Applications in Medicine
How can data science help doctors and clinicians treat patients better?

- Week 9 Module 9: Q2-1 Electronic Medical Records, Clinical Trials, Precision Medicine
- Week 10 Module 10: Q2-2 Apps for Doctors, Nurses, Clinicians
- Week 11 Module 11: Q2-3 Clinician Perspective (Grant Cumming), Government, Public Health

Data Science Applications in Personal Health
And how can data science help us all lead healthier lives?

- Week 12 Module 12: Q3-1 Patient Focused. (Quantified Self, Electronic Diaries, Your Medical Mind)
- Week 13 Module 13: Q3-2 Crowdsourcing, online communities, ePatients, patient advocates
- Week 14 Module 14: Q3-3 Patient Perspective (Grant Cumming) Personal Genomics

Virtual Poster Session – Student Showcase and Course Wrap-up
- Week 15 Module 15: Virtual Poster Session (presentation of student projects)
How can Data Science help researchers in drug discovery?

How can Data Science help doctors and other providers?

How can Data Science help us all lead healthier lives?