DATA SCIENCE (DATASCI)

DATASCI 202 Opportunities and challenges of complex biomedical data (3 Units)

Offered In: Summer

Instructor(s): Karla Lindquist
Prerequisite(s): None

Restrictions: This course is part of the Health Data Science Program and may have space limitations. Auditing is not permitted.

Activities: Lecture, Project, Lab skills

This is an introduction to the opportunities and challenges of using large datasets for biomedical research. Topics to be covered include: What makes big data different? What big data can and cannot do. Phases of data science: getting data, merging and cleaning data, storing and accessing data, visualizing or telling stories with data, drawing conclusions from data. Introduction to supervised and unsupervised machine learning including detailed discussion of algorithms and model fitting.

School: Graduate Division
Department: Health Data Science Program
May the student choose the instructor for this course? No
Does enrollment in this course require instructor approval? No
Course Grading Convention: Letter Grade, P/NP (Pass/Not Pass) or S/U (Satisfactory/Unsatisfactory)
Graduate Division course: Yes
Is this a web-based online course? No
Is this an Interprofessional Education (IPE) course? No
May students in the Graduate Division (i.e. pursuing Master or PhD) enroll in this course? Yes
Repeat course for credit? No

DATASCI 213 Programming for Health Data Science in R (2 Units)

Offered In: Summer

Instructor(s): Stathis Gennatas
Prerequisite(s): No prior programming experience is required.

Restrictions: This course is part of the Training in Clinical Research (TICR) and Health Data Science Program and may have space limitations. Auditing is not permitted.

Activities: Lecture, Lab science

Vast amounts of health-related data are being generated daily and at an increasing rate. Our ability to extract insights and make the most of these resources depends on the effective and efficient use of computational tools to preprocess, visualize, and analyze different types of data. This course is an introductory programming course which aims to provide hands-on experience in the R language and enable further work in biostatistics, epidemiology, and machine learning/health data science.

School: Graduate Division
Department: Health Data Science Program
May the student choose the instructor for this course? No
Does enrollment in this course require instructor approval? No
Course Grading Convention: Letter Grade, P/NP (Pass/Not Pass) or S/U (Satisfactory/Unsatisfactory)
Graduate Division course: Yes
Is this a web-based online course? No
Is this an Interprofessional Education (IPE) course? No
May students in the Graduate Division (i.e. pursuing Master or PhD) enroll in this course? Yes
Repeat course for credit? No

DATASCI 214 Programming for Health Data Science in R II (2-3 Units)

Offered In: Fall

Instructor(s): Stathis Gennatas, John Kornak
Prerequisite(s): DATASCI 213 or equivalent.

Restrictions: This is a core course of the Health Data Science (HDS) program and part of the Training in Clinical Research Program and may have space limitations. Auditing is not permitted.

Activities: Lecture, Lab skills

R programming course to enable work in any field including biostatistics, epidemiology, data science/machine learning. This course builds on students prerequisite core R language knowledge to cover skills in advanced data transformations, visualization, working with big (in-memory) data, report-writing, and core statistic testing.

School: Graduate Division
Department: Health Data Science Program
May the student choose the instructor for this course? No
Does enrollment in this course require instructor approval? No
Course Grading Convention: Letter Grade, P/NP (Pass/Not Pass) or S/U (Satisfactory/Unsatisfactory)
Graduate Division course: Yes
Is this a web-based online course? No
Is this an Interprofessional Education (IPE) course? No
May students in the Graduate Division (i.e. pursuing Master or PhD) enroll in this course? Yes
Repeat course for credit? No

DATASCI 216 Machine Learning in R for the Biomedical Sciences (3 Units)

Offered In: Winter

Instructor(s): Adam Olshen
Prerequisite(s): BIOSTAT 208, DATASCI 213 & BIOSTAT 209. Exceptions to these prerequisites may be made with the consent of the Course Director, space permitting. Strongly recommended: EPIDEMIOL 204 & DATASCI 202

Restrictions: This is a core course of the Health Data Science (HDS) program and part of the Training in Clinical Research Program and may have space limitations. Auditing is not permitted.

Activities: Lecture, Project

This is a course that covers machine learning methods as they apply to areas of biomedical research and will teach how to implement the methods in R. Topics to be covered include: What is Machine learning? Prediction techniques (including classification) and methods for assessing them, Cross-validation, penalized regression methods such as lasso, boosting, bagging and ensemble methods, pattern recognition, deep learning, and data reduction methods, and machine learning meta packages in R.

School: Graduate Division
Department: Health Data Science Program
May the student choose the instructor for this course? No
Does enrollment in this course require instructor approval? No
Course Grading Convention: Letter Grade, P/NP (Pass/Not Pass) or S/U (Satisfactory/Unsatisfactory)
Graduate Division course: Yes
Is this a web-based online course? No
Is this an Interprofessional Education (IPE) course? No
May students in the Graduate Division (i.e. pursuing Master or PhD) enroll in this course? Yes
Repeat course for credit? No

DATASCI 217 Introduction to Python and Data Science Tools (1-2 Units)

Offered In: Fall

Instructor(s): John Kornak
Prerequisite(s): BIOSTAT 213 or equivalent (knowledge of probability/statistics and familiarity with programming concepts, e.g., from using R)

Restrictions: This course is part of the Health Data Science Masters and Certificate Program and may have space limitations.

Activities: Lecture, Workshop

This course provides an introduction to essential tools and skills for data science, focusing on Python programming and industry-relevant tools. Students will learn command line basics, version control with Git, documentation with Markdown, remote execution, and high-performance computing (HPC). Integrated throughout the course, the Python component covers syntax, flow control, data management, visualization, libraries for data science, and algorithms and data structures common in interviews.

School: Graduate Division
Department: Health Data Science Program
May the student choose the instructor for this course? No
Does enrollment in this course require instructor approval? No
Course Grading Convention: Letter Grade, P/NP (Pass/Not Pass) or S/U (Satisfactory/Unsatisfactory)
Graduate Division course: Yes
Is this a web-based online course? No
Is this an Interprofessional Education (IPE) course? No
May students in the Graduate Division (i.e. pursuing Master or PhD) enroll in this course? Yes
Repeat course for credit? No

DATASCI 220 Data Science Program Seminar I (1 Units)

Offered In: Fall, Winter, Spring

Instructor(s): John Kornak
Prerequisite(s): BIOSTAT 202 and BIOSTAT 213

Restrictions: This course is restricted to students enrolled in the Certificate in Health Data Science and the Master's degree in Health Data Science (first year students).

Activities: Seminar, Independent Study

This seminar series covers topics in data science algorithms, ethics, biases, and applications. Students will be exposed to current topics on Data Science and Machine Learning/Biostatistics and Health Data applications, discuss issues in data science, present their work, and learn how to critically evaluate research literature. External speakers will be invited to give presentations on potential careers in health data science across the biotech industry, government and academia.

School: Graduate Division
Department: Health Data Science Program
May the student choose the instructor for this course? No
Does enrollment in this course require instructor approval? No
Course Grading Convention: P/NP (Pass/Not Pass) or S/U (Satisfactory/Unsatisfactory)
Graduate Division course: Yes
Is this a web-based online course? No
Is this an Interprofessional Education (IPE) course? No
May students in the Graduate Division (i.e. pursuing Master or PhD) enroll in this course? Yes
Repeat course for credit? Yes

DATASCI 221 Data Science Program Seminar II (1 Units)

Offered In: Fall, Winter, Spring

Instructor(s): John Kornak
Prerequisite(s): DATASCI 220

Restrictions: This course is restricted to students enrolled in year 2 of the Master's in Health Data Science program.

Activities: Seminar, Independent Study

This course covers advanced topics of data science methods, ethics and biases. The focus in this second year of the seminar program will be on students presenting their research work progress from their Capstone projects. Additionally, students will also learn how to critically evaluate research literature.

School: Graduate Division
Department: Health Data Science Program
May the student choose the instructor for this course? No
Does enrollment in this course require instructor approval? No
Course Grading Convention: P/NP (Pass/Not Pass) or S/U (Satisfactory/Unsatisfactory)
Graduate Division course: Yes
Is this a web-based online course? No
Is this an Interprofessional Education (IPE) course? No
May students in the Graduate Division (i.e. pursuing Master or PhD) enroll in this course? Yes
Repeat course for credit? Yes

DATASCI 222 Data Science Capstone Project (8 Units)

Offered In: Fall, Winter, Spring

Instructor(s): John Kornak
Prerequisite(s): BIOSTAT 202, BIOSTAT 213, BIOSTAT 214, BIOSTAT 216, DATASCI 220, DATASCI 225

Restrictions: This course is restricted to 2nd year students in the Master's in Health Data Science program.

Activities: Project

Capstone project requirement for students in the Masters in Health Data Science program. Students will write a first author paper researching a problem in health data science and analyzing data using appropriate data science methodology; present their work at a scientific conference; generate a portfolio of code, analyses and data products; and write a detailed report on the background methodology and technical issues that were considered as well as implemented for the submitted publication.

School: Graduate Division
Department: Health Data Science Program
May the student choose the instructor for this course? No
Does enrollment in this course require instructor approval? No
Course Grading Convention: P/NP (Pass/Not Pass) or S/U (Satisfactory/Unsatisfactory)
Graduate Division course: Yes
Is this a web-based online course? No
Is this an Interprofessional Education (IPE) course? No
May students in the Graduate Division (i.e. pursuing Master or PhD) enroll in this course? Yes
Repeat course for credit? Yes

DATASCI 223 Applied Data Science with Python (2 Units)

Offered In: Spring

Instructor(s): John Kornak
Prerequisite(s): Familiarity with programming concepts, including loops, variables, and functions. Ideally, hands-on experience writing and running scripts such as in: Python, R, Bash, or other programming languages.

Restrictions: This course is part of the Health Data Science Masters and Certificate Program and may have space limitations. Auditing is not permitted.

Activities: Lecture, Project, Workshop

Survey of Data Science methods in Python, starting with common data science tools and processes and spending one week per topics learning to build common ML/AI solutions.

School: Graduate Division
Department: Health Data Science Program
May the student choose the instructor for this course? No
Does enrollment in this course require instructor approval? No
Course Grading Convention: Letter Grade, P/NP (Pass/Not Pass) or S/U (Satisfactory/Unsatisfactory)
Graduate Division course: Yes
Is this a web-based online course? No
Is this an Interprofessional Education (IPE) course? No
May students in the Graduate Division (i.e. pursuing Master or PhD) enroll in this course? Yes
Repeat course for credit? No

DATASCI 224 Understanding Machine Learning: From Theory to Applications (3 Units)

Offered In: Spring

Instructor(s): Jean Feng
Prerequisite(s): BIOSTAT 216

Restrictions: This course is part of the Health Data Science Masters and Certificate Program and may have space limitations. Auditing is not permitted.

Activities: Lecture, Project

This course teaches the mathematical foundations of machine learning (ML). Each week, the course surveys a different algorithm to examine its underlying machinery, covering topics such as linear algebra, calculus, and optimization. ML algorithms range from linear models to gradient boosting and deep learning. The course also discusses newer concepts such as model fairness and ML for causal inference. Upon course completion, students should be able to learn new ML algorithms independently.

School: Graduate Division
Department: Health Data Science Program
May the student choose the instructor for this course? No
Does enrollment in this course require instructor approval? No
Course Grading Convention: Letter Grade, P/NP (Pass/Not Pass) or S/U (Satisfactory/Unsatisfactory)
Graduate Division course: Yes
Is this a web-based online course? No
Is this an Interprofessional Education (IPE) course? No
May students in the Graduate Division (i.e. pursuing Master or PhD) enroll in this course? Yes
Repeat course for credit? No

DATASCI 226 Bayesian Methods and Gaussian Processes (2-3 Units)

Offered In: Fall

Instructor(s): John Kornak
Prerequisite(s): Basic knowledge of probability and statistics (BIOSTAT 200 and BIOSTAT 208 equivalent); programming skills in R (BIOSTAT 213 and BIOSTAT 214 equivalent); some familiarity with calculus and linear algebra (especially for the extra Gaussian processes unit).

Restrictions: This course is part of the Health Data Science Masters and Certificate Program and may have space limitations. Auditing is not permitted.

Activities: Lecture, Project

This course provides an introduction to Bayesian statistics, Markov Chain Monte Carlo (MCMC) sampling, and Gaussian Processes. The first two units cover the fundamentals of Bayesian methods and MCMC, and the final optional unit explores Gaussian processes. Students will gain practical skills in applying these techniques to real-world problems using R, STAN, and JAGS.

School: Graduate Division
Department: Health Data Science Program
May the student choose the instructor for this course? No
Does enrollment in this course require instructor approval? No
Course Grading Convention: Letter Grade, P/NP (Pass/Not Pass) or S/U (Satisfactory/Unsatisfactory)
Graduate Division course: Yes
Is this a web-based online course? No
Is this an Interprofessional Education (IPE) course? No
May students in the Graduate Division (i.e. pursuing Master or PhD) enroll in this course? Yes
Repeat course for credit? No

DATASCI 300 Data Science Educational Practice (2 Units)

Offered In: Fall, Winter, Spring, Summer

Instructor(s): Staff
Prerequisite(s): Students must have previously taken the course they EA for.

Restrictions: This course is restricted to 2nd year students in the Master's in Health Data Science program.

Activities: Lab science, Discussion

Masters in Health Data Science students are expected to act as an educational apprentice (EA). This experience involves leading a weekly small-group discussion section of 10-15 students, holding office hours and grading homework assignments and projects. This requirement will provide students with valuable teaching experience without having a significant time impact on their Capstone project work. In all cases, students will have taken the course they are asked to EA during their first year.

School: Graduate Division
Department: Health Data Science Program
May the student choose the instructor for this course? Yes
Does enrollment in this course require instructor approval? No
Course Grading Convention: P/NP (Pass/Not Pass) or S/U (Satisfactory/Unsatisfactory)
Graduate Division course: Yes
Is this a web-based online course? No
Is this an Interprofessional Education (IPE) course? No
May students in the Graduate Division (i.e. pursuing Master or PhD) enroll in this course? Yes
Repeat course for credit? No