USC Viterbi
School of Engineering

ISE 503: Health Analytics

4 Units
Day/Time: 110-minute twice weekly
Location: TBD

Instructor: Abigail Horn
Office:

Office Hours: TBD (in person)
Virtual office hours by appointment

Contact Info: Abigail.Horn@usc.edu

Teaching Assistants:
TBD

Catalog Course Description

Case-study approach to learning descriptive and predictive data analytical tools for health data. Derive inspiring solutions to real-world problems in various health domain applications.

Expanded Course Description

Health decisions can have life-altering outcomes for patients and populations. The increased ability to collect health-related data, combined with innovations in data analytics techniques, are revolutionizing the ability of decision-makers to make evidence-based choices that can dramatically improve patient and population health, well-being, and experience, and impact healthcare costs. Health data analytics, at the intersection of statistics, programming, and healthcare, provides a foundation for transforming data into insights and predictions to inform sound, impactful decision-making for problems across many health-domain problems.

Through a survey of health analytics case studies, this course will overview different types of health domain applications, health data, and descriptive and predictive analytical techniques available to address health-domain problems, using the R programming language. The primary objective will be to provide students with the skills and experience to implement existing data analytics techniques to gain insights and make predictions with health data, to interpret results to determine possible decisions and their advantages and disadvantages, and to communicate these findings in the context of achieving results in health.

The course is intended for M.S. students in an analytics or data science master's program who are interested in the health domain, or a health sciences master's program and have experience in data analytics.

Learning Objectives

The overall course objective is to learn to identify, apply, and interpret appropriate analytic methods and health data types to address real-world health-domain problems. Upon successful completion of this class, the student will be able to

- Identify health data types including electronic medical records, medical text, medical claims data, disease registry data, healthcare survey data, and social and behavioral data.
- Describe the heterogeneity in these data sources and the diversity of analytical techniques necessary to approach them.
- Classify existing descriptive and predictive analytics techniques and algorithms to gain insights and make predictions with health data including data exploration and visualization, linear regression and extensions, logistic regression, decision and classification trees, random forests, clustering, text analytics, and causal inference.

- Interpret results from the application of these existing analytics methods to determine possible decisions, to identify the advantages and disadvantages of these directions in terms of their health impacts, and to communicate these findings to health sector decision-makers.

Prerequisite(s):

None.

Recommended Preparation:

Statistics on the level of ISE 225; programming on the level of ISE 150 (familiarity with R would be helpful). This course will illustrate analytics concepts taught in ISE-529 and ISE-535 using examples from health. Previously or concurrently taking these courses, especially ISE-529 (Predictive Analytics) is extremely helpful, but is not required.

Course Overview

The beginning of the course will provide an introduction and overview of health domain sectors and challenges, heath domain data, and analytics methods that can be applied to address challenges using this data. The remainder of the course is structured into a series of case studies, each addressing a different health domain problem, health domain data type, and data analytics method. Case studies will be analyzed during back-to-back lecture and lab components of the class. The lecture component will involve a conceptual introduction to the methods followed by demonstrations of how to apply the methods to analyze the data. This will provide all of the content the students need to implement the analytics method being developed. Class discussion will be emphasized throughout the lecture. The lab session students will provide students with hands-on experience working with the data to replicate the implementation of the methods demonstrated during the lecture, and to attempt extensions to these solutions.

The course will involve a team project of a topic of students' choosing relating to health analytics (details below) which will begin half-way through the semester and end with final reports and a short team-to-team interview-based presentation. The project is a significant aspect of the class, as reflected in the grading scheme.

The course will also involve assignments designed to develop interpretation of health analytics concepts and their implementation in practice to drive health results, through a critique of a paper or report, and a summary of a guest lecture.

Technological Proficiency and Hardware/Software Required

The R programming language and R Studio will be used in most class demonstrations and assignments. Python will be used in one class demonstration. These software are freely available.

Required Readings and Supplementary Materials

All required course content will be included in the lecture notes. We will be posting research papers of significance to particular health analytics topics to complement the materials covered in class via Brightspace. To go deeper or for additional reference, you may find it useful to refer to the following textbook:

- James, et. al., *An Introduction to Statistical Learning with Applications in R,* 2nd edition, Springer, 2021 (ISLR2) (this book is available for free download on the authors' website at: https://www.statlearning.com/)

Description and Assessment of Assignments

Homework Assignments
There will be 7 homework assignments given throughout the semester, approximately 1 every 2 weeks. These will aim to reinforce your understanding of the methods covered in class and their computational implementation, and interpretation of results. Some assignments will also cover broader data visualization, problem framing, and data-driven communication skills.

All homework assignments are individual. You may find it useful to discuss the problems with one another, however individual solutions must be submitted and copying will not be tolerated.

In-class Labs
Submission of labs conducted in class is mandatory. Labs must be completed in R Markdown, and submitted in PDF form using Gradescope. Labs will be graded for completeness and clarity rather than accuracy.

Paper Critique
Several research papers showcasing recent health domain research using data analytics methods will be shared. Students will choose one paper and write a short critique. The critique should (i) outline the health domain application problem and explain why it is important; (ii) summarize the analytics method(s) used and the results; (iii) interpret the findings in the context of who/what they apply to (patients, population health, insurers, etc.); and (iv) discuss limitations of the approaches and suggest avenues for future work. Your critique should be a maximum of 1 page long.

Final Project
This will be a project-based class. The project will provide the opportunity to identify a realistic health challenge and to apply analytics methods to address the problem. You will work in self-formed pairs, teams of two. Each team will choose a health application area and identify the decision-making problem, gather the relevant data, use analytics to conduct the analysis, and interpret and communicate findings and recommendations. Findings will be communicated in the form of a report and an interview-based presentation to a fictitious health sector client. While the choice of health application area is flexible, each team should be able to clearly articulate how the application and decision-making problem relate to the health sector. The project will reinforce both technical skills (e.g., programming, model development, model interpretation, data visualization and reporting) and "soft" skills that are very important for working in the health industry (e.g., teamwork, collaboration, communication, project management). Sample project topics have included Breast Cancer Survival Prediction, Medical Student Mental Health Status Prediction, Health Insurance Cost Prediction, Key Indicators of Myocardial Infarction, and Factors Contributing to Car Crash Fatality.

Deliverables throughout the semester will guide students through conducting the project (dates TBD):

- **By Week 6**: Each team needs to be registered with a topic. (5% of the project grade)
- **By Week 10**: Each team submits an interim report (1-2 pages, excluding appendices) (20% of project grade).
    - The report should describe the health application area, the decision-making problem, the data, the analytics approach, initial results, the next steps, and the expected impact.
- **By Week 14**: Each team submits the following deliverables (all in pdf):
    - An executive summary including recommendations written for a senior decision-makers (around 300 words) (15% of project grade).
    - A report (around 4 pages, excluding appendices) that includes a description of the health application area, the problem, your analysis, your results, and implications for health or health industry (30% of project grade).
- **On Week 15**: We will hold interviews to showcase your work (30% of project grade). Each team will be interviewed by another team. The presenting team will first be given an opportunity to present its work

(2–3 minutes, 2–3 slides). The interviewing team will play the role of a key decision-maker, in charge of taking and implementing the recommendations.

Participation

Class participation will be evaluated based on engagement in class discussion. Meaningful engagement may include participation in discussion Q&A (asking or answering questions from the instructor or other students), asking or answering questions during lecture, and engaging in guest lectures. For students who miss lecture, a 0.5-1 page summary of the key concepts taught in the class and the key points from the in-class discussion can be contributed after watching the lecture video. At least 5 meaningful class interactions are needed for full participation points.

Grading Breakdown

Assignment	% of Grade
Homework Assignments (6)	30
Lab Assignments (12)	15
Paper Critique	5
Guest Speaker Summary	5
Final Project	40
Class Participation	5

Assignment Submission Policy

Assignments will be posted on Brightspace and submitted using GradeScope, a grading system that allows for detailed feedback (instructions will be provided).

You will be given 2 weeks for most assignments. Please refer to the assignment posting on Blackboard for the exact length and due date. Please make sure to submit on time. Unless otherwise noted, assignments turned in after the due date will be penalized by 25%. Assignments submitted >48 hours of due date will not be accepted.

Course Schedule

The broad outline of topics and timelines is summarized below.

The course will focus on descriptive and predictive analytics, focusing on both key methods and modern developments:

Descriptive and predictive analytics (Sessions 1 – 9): These sessions will focus on the key analytics methods used to extract insights and predictions from health data, including: linear and non-linear regression, model selection, logistic regression, support vector machines, classification and regression trees, and random forests.

Modern analytics developments (Sessions 10 – 13): These sessions will focus on health applications of recent analytics developments from novel techniques or existing techniques that are now receiving increased attention in the machine learning and analytics communities.

Introduction

- Introduction to the 3 dimensions we will explore in this course: health domain applications, health data types, and data analytics approaches
- Introduction to R for data analysis

Case-study lectures demonstrating application in interpretation of the following data analytics methods (see description of this class format under Course Overview above):

1. Exploratory data analysis and data visualization
2. Linear regression
3. Non-linear regression
4. Regularization
5. Resampling and cross-validation
6. Logistic regression
7. Regression trees (CART)
8. Classification trees (CART)
9. Random forests
10. Survival analysis
11. Deep learning
12. Causal inference
13. Guest lecture

Closure

- Final Project Interviews (1 week)

A detailed week-by-week course breakdown is found on the following pages.

Please note that the schedule may change as we go. Please refer to Brightspace and class announcements for the latest schedule.

Week	Method	Case	Work Assigned	Work Due
1	**Introduction to R:** Brief introduction including RStudio, rmarkdown	**Introduction to 3 dimensions of this course:** Health data types and sources, Health domain problems, Data analytics methods	**HW0** (graded for completeness)	
Descriptive and Predictive Analytics				
1 cont'd	**Exploratory data analysis (EDA) and data visualization**	**Objective**: Explore patterns in relationships between environmental factors and air pollution on children's respiratory health **Purpose**: Inform hypothesis generation, relationships to explore in estimation and prediction models **Data types**: Health survey (*The USC Children's Health Study*), spatial data	**HW1**: EDA	**HW0**

Week	Method	Case	Work Assigned	Work Due
2	**Linear regression**	**Objective**: Predict children's forced expiratory volume (FEV) from routine clinical characteristics, compare to observed value **Purpose**: Clinical tool for diagnosing lung disease **Data types**: [*Repeated data*] Health survey (*The USC Children's Health Study*), spatial data		
3	**Non-linear regression**	**Objective**: Forecast pm2.5 level by neighborhood **Purpose**: Support hospital planning for acute cardiovascular events **Data types**: Spatial, environmental	**HW2:** Linear regression, regularization, cross validation	**HW1**
4	**Linear regression with regularization**	**Objective**: Predict future diabetes from medical records **Purpose**: Identify patients to prioritize for early treatment **Data types**: Electronic medical records (demographic characteristics, clinical measurements, and lab results)		
5	**Resampling and cross-validation**	**Objective**: [*Repeated case*] Predict heart disease diagnosis **Purpose**: Identify patients who should be prescribed preventive medication **Data types**: Health survey data (*Framingham Heart Study*)	**Final Project:** Teams register with topic	**HW2**
6	**Logistic regression**	**Objective**: Predict future heart disease diagnosis **Purpose**: Identify patients who should be prescribed preventive medication **Data types**: Health survey data (*Framingham Heart Study*)	**HW3**: Logistic regression	
7	**Regression trees (CART)**	**Objective**: Predict future expenditures from healthcare claims data **Purpose**: Identify high-risk patients to mitigate future risks **Data types**: Healthcare claims data (demographics, expenditures, eligibility, diagnoses, procedures)		

Week	Method	Case	Work Assigned	Work Due
8	**Classification trees (CART)**	**Objective**: Identify neighborhood predictors of neighborhood-level prevalence of stroke **Purpose**: Prioritize communities for stroke-related policy interventions **Data types**: Spatial data (*500 Cities Project*); administrative data (*U.S. Census*)	**HW4**: CART	**HW3**
9	**Random forests**	**Objective**: [*Repeat case*] Predict future expenditures from healthcare claims data **Purpose**: Identify high-risk patients to mitigate future risks **Data types**: Healthcare claims data (demographics, expenditures, eligibility, diagnoses, procedures)	**Paper Critique**	**HW4**
Modern Analytics Developments				
10	**Survival analysis and algorithmic fairness**	**Objective**: Assess fairness of predictive model of recidivism following parole **Purpose**: Illuminate tradeoffs among fairness criteria, and guide model selection accordingly **Data types**: Administrative data (demographics, criminal history)	**HW5**: Survival analysis	
11	**Deep Learning**	**Objective**: TBD **Data types**: TBD		**Final project**: Interim report
12	**Causal inference** ***Guest Lecture***	**Objective**: Understand differences in requirements and use of causal models vs. estimation and prediction models; become familiar with several methodologies for causal reasoning inference developed within computer science **Purpose**: Be able to apply models for future reasoning **Data types**: TBD	**HW6**: Causal Inference	**HW5**

Week	Method	Case	Work Assigned	Work Due
13	**Causal inference** ***Guest Lecture***	**Objective**: Understand differences in requirements and use of causal models vs. estimation and prediction models; become familiar with several methodologies for causal reasoning inference developed within computer science **Purpose**: Be able to apply models for future reasoning **Data types**: TBD		
14	***Guest Lectures***		**HW 7**: Guest speaker summary	**HW6**
15		**Final project presentations and interviews**		**Final project** (executive summary due 3 days before presentation)
FINAL		**Final project** (executive summary)	Refer to the final exam schedule in the USC Schedule of Classes at www.usc.edu/soc	

Statement on Academic Conduct and Support Systems

Academic Integrity

Unless otherwise noted, this course will follow the expectations for academic integrity as stated in the USC Student Handbook. The general USC guidelines on Academic Integrity and Course Content Distribution are provided in the subsequent "Statement on Academic Conduct and Support Systems" section.

In particular, for this class: Besides the Final Project, which is specifically designated as a 'group project,' all assignments are expected to be completed individually. This includes homework and lab assignments. You are welcome to discuss problems, approaches, and solution strategies with your peers. However, individual solutions must be submitted and copying / plagiarism will not be tolerated.

Plagiarism includes the submission of writing or code written by, or otherwise obtained from another source, including generative AI (see section below on allowed and disallowed use of generative AI in this course).

Other violations of academic integrity include, but are not limited to, cheating, fabrication (e.g., falsifying data), collusion, knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university. All incidences of academic misconduct will be reported to the Office of Academic Integrity and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see the student handbook or the Office of Academic Integrity's website, and university policies on Research and Scholarship Misconduct.

Please ask the instructor [and/or TA(s)] if you are unsure about what constitutes unauthorized assistance on an exam or assignment, or what information requires citation and/or attribution.

Use of Generative AI in this Course

Use of generative AI (i.e., large language models – LLMs – like ChatGPT) is permitted but limited as follows: In this course, you are permitted to use artificial intelligence (AI)-powered programs to help you, but only on assignments that explicitly indicate a permitted use of AI. The situations where use of generative AI will NOT be allowed include:

- Helping to generate content in lab or homework assignment involving (i) the creation of code, and (ii) interpretation of the health context of a problem. Developing and using analytical and critical thinking skills to solve problems and answer health-related questions are part of the learning outcomes of this course. Therefore, assignments should be prepared by the student working individually or in groups as described on each assignment
- Generating slides or drafting narrative text for your final project presentation
- Writing your final project report

Situations where use of generative AI WILL be allowed include:

- Helping to understand content provided in lectures
- Helping brainstorm topics for your final project
- Helping brainstorm important background context to include in the introduction section of your final report
- Revising drafts of a report you have already written
- Helping to understand content provided in lectures

When you do choose to use generative AI, please keep in mind:

- Be aware that AI text generation tools may present incorrect information, biased responses, and incomplete analyses. If it gives you reference, fact, or number: assume it is incorrect unless you either know the correct answer or can verify its accuracy with another source. You work will be evaluated for correctness and bias, and references will be checked. You will therefore be responsible for any errors or omissions provided by the tool.
- Additionally, it is often very easy to spot out-of-the-box uses of AI-generated text. Your work will be evaluated for originality, so it is important to make any work you submit your own through revision.
- If you provide minimum-effort prompts, you will get low-quality results. You will need to refine your prompts to get good outcomes. This will take multiple iterations.
- AI is a tool, but one that you need to acknowledge using. Please *include a paragraph at the end of any assignment explaining if, how, and why you used AI and indicate/specify the prompts you used to obtain the results*. If using generative AI to help in any aspect of your final project, you will need cite it. Including this text will not count towards page limits for an assignment, if given. Please credit AI-generated material in your final project in the References section; this link from USC Libraries explains how. Failure to include this is a violation of academic integrity policies.

The above is a general list of times when and when not to use AI in course materials. Please r review the instructions in each assignment for more details on how and when to use AI Generators for your submissions.

Course Content Distribution and Synchronous Session Recordings Policies

You may not record this class without the express permission of the instructor and all other students in the class. Distribution of any notes, recordings, exams, or other materials from a university class or lectures — other than for individual or class group study — is prohibited without the express permission of the instructor.

USC has policies that prohibit recording and distribution of any synchronous and asynchronous course content outside of the learning environment.

Recording a university class without the express permission of the instructor and announcement to the class, or unless conducted pursuant to an Office of Student Accessibility Services (OSAS) accommodation. Recording can inhibit free discussion in the future, and thus infringe on the academic freedom of other students as well as the instructor. (Living our Unifying Values: The USC Student Handbook, page 13).

Distribution or use of notes, recordings, exams, or other intellectual property, based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study. This includes but is not limited to providing materials for distribution by services publishing course materials. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in relationship to the class, whether obtained in class, via email, on the internet, or via any other media. (Living our Unifying Values: The USC Student Handbook, page 13).

Students and Disability Accommodations:
USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Support Systems:

Counseling and Mental Health - (213) 740-9355 – 24/7 on call
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

988 Suicide and Crisis Lifeline - 988 for both calls and text messages – 24/7 on call
The 988 Suicide and Crisis Lifeline (formerly known as the National Suicide Prevention Lifeline) provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week, across the United States. The Lifeline is comprised of a national network of over 200 local crisis centers, combining custom local care and resources with national standards and best practices. The new, shorter phone number makes it easier for people to remember and access mental health crisis services (though the previous 1 (800) 273-8255 number will continue to function indefinitely) and represents a continued commitment to those in crisis.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL) – 24/7 on call
Free and confidential therapy services, workshops, and training for situations related to gender- and power-based harm (including sexual assault, intimate partner violence, and stalking).

Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - (213) 740-5086
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

The Office of Student Accessibility Services (OSAS) - (213) 740-0776
OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

USC Campus Support and Intervention - (213) 740-0411
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity, Equity and Inclusion - (213) 740-2101
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-1200 – 24/7 on call
Non-emergency assistance or information.

Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)
A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

Occupational Therapy Faculty Practice - (323) 442-2850 or otfp@med.usc.edu
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.