



This syllabus is currently awaiting approval. Please note that the syllabus preview and any edits may not yet be published.



Fall Term 2023 - Full Term · CSCI E-96 1 ·

Refresh
Edit syllabus



Harvard Extension School

HARVARD DIVISION OF CONTINUING EDUCATION

Course Learning Outcomes

Mode of Attendance & Participation Policy

Grading & Grade Definitions

Course Materials

CSCI E-96

Academic Integrity Policy

Data Mining for Business

Accessibility Services Policy

Publishing or Distributing Course Materials

Fall Term 2023 Policy

Class Meeting Schedule

Course Information

CRN: 15736

Section Number: 1

Format: Flexible Attendance Web Conference

Credit Status: Graduate

Credit Hours: 4

Class Meetings: Mondays, September 11-December 2
10:10pm

Course Description: This course introduces non-math business professionals to data science principles widely used in today's corporations. Quantitative methods affect many business interactions for business leaders, students, and consumers. Emphasis is placed on practical uses and case studies of data to inform business decisions rather than theoretical or complex mathematics. Case study topics include understanding customer demand, marketing, new market forecasting, projections, and data mining to improve decisions. Learners will include quantitative business application, basic program

algorithm development, and process workflow. The course highlights methods that business leaders and data scientists found to be the most useful. It introduces the basic concepts for data mining. This course is for students who want an introduction to how data science improves business outcomes.

Prerequisites: Since this course utilizes R throughout the semester, students should complete the 4-hour free online course *Introduction to R* at DataCamp.com found here: <https://www.datacamp.com/courses/free-introduction-to-r>

Grading & Grade Definitions
Course Materials
Academic Integrity Policy
Instructor Information & Office Hours

Publishing or Distributing Course Materials
Ted Kwartler
Email: edwardkwartler@fas.harvard.edu

Kalle Georgiev
Email: kgeorgiev@g.harvard.edu

Course Goals / Learning Outcomes

If you stay engaged in the course and complete the suggested readings and assignments:

You will be able to think systematically about how data make business decisions. This objective will be accomplished through the use of ideas from statistics, economics, and technology and using business-related case studies.

Students will learn how to implement a variety of popular mining algorithms in R (a free and open-source software) to solve business problems and identify opportunities. This course introduces the basics of R in data mining.

As a business leader, you will acquire the skill of applying science concepts within business domains to improve and learn how data scientists approach projects.

As a data scientist, you will acquire practical application mining methods that are used in many of today's most organizations as well as understand what business stakeholders expect of data scientists.



Mode of Attendance & Participation Policy

This class offers a live or on-demand option, which means you can choose to attend the class live over Zoom or watch the recording afterward. You do not need to commit to the mode of attendance for the whole semester.

If you are attending live over Zoom:

Class meetings take place over Zoom. Because they involve active participation, discussion, and dialogue, you are expected to attend all class meetings. Please arrive on time. You should join Zoom meetings with a functional web-camera and microphone. Please be prepared with materials needed, to engage thoughtfully in discussion. Please keep your camera on. You may turn off your camera for occasional interruptions or momentarily for privacy.

You will also need the most up-to-date Zoom client installed on your computer to join class. Please participate from a suitable and appropriate environment with appropriate clothing for class. Participating while traveling or in a car is not permitted. Please do not join class via mobile phone or web browser.

If you are participating on demand:

You are expected to watch the class recording and complete assignments before the next live class meets.

Please be sure to review important information on [Study Policies and Conduct](#).



Grading & Grade Definitions

Grading

A course grade will be assigned based on student performance on case studies, applicable homework assignments, and assignment.

Course Information

Instructor Information & Office Hours

Course Sign-Up Based on Student

Mode of Attendance & Participation Policy

Grading & Grade Definitions

Course Materials

Academic Integrity Policy

Any


Class Meeting Schedule

Assignments are accepted up to 12 hours late. Any work submitted after the deadline but before 12 additional hours is penalized 10% of the total weight of the assignment. After 12 hours, no late submissions will be accepted under ANY circumstances. Pupils are expected to manage their own time and submit work accordingly. Failure to submit submissions through the University approved portal by the assignment deadline is considered late and not accepted. Submissions to any other location will not be accepted.

Graduate Student Grading (4 cases)

- Skills Assessment: 0%: Complete the provided but unfinished R script.
- Case I 25% of final grade: EDA Case
- Case II 25% of the final grade: Banking Case
- AI Ethics Case 25%: Build a model evaluating its accuracy and unfair bias of a protected feature
- Extra Credit: Homework II Visualization in R 1% of

Undergraduate Student Grading (only 3 cases)

- Skills Assessment: 0%: Complete the provided but unfinished R script.
- Case I 25% of final grade: EDA Case  Refresh
- Case II 25% of the final grade: Banking Case
- Homework I: 10% Intro to R script - more info to c
- Homework II: 20% Visualization in R script - more come
- Homework III: 20% Obtain 2 AI/Data Ethics relate (https://incidentdatabase.ai is a good resource):
 - Use ChatGPT to summarize the article
 - Critique the summarization as appropriate, inappropriate, missing relevant facts, creating information from outside the article etc. in a paragraph
 - Write 1 paragraph WITHOUT GPT with your reflection on the use or misuse of the technique in the article. In the paragraph suggest way mitigate or monitor to protect against the issue the article.

Grading Scale

You earn the grade based on assignments according to below. Grades are not curved to fit a predetermined distribution. A student's degree, certificate candidacy, or funding status have any impact on a course grade. "Needing an A" for any reason is not sufficient to earn an A grade. Note there are "minus" grades given in the course. It is the belief of the department that minus grades constitute a false precision in many cases and further penalize frequent "A-" students since there is no way to obtain an "A+" to rebalance a GPA. To the student

benefit, one can still earn a “plus” on their final grade at the scale below.

Max	Min	Grade
100	90	A
89.9	87	B+
86.9	80	B
79.9	77	C+
76.9	70	C
69.9	67	D+
66.9	60	D
59.9	0	F



Course Information
 Instructor Information & Office Hours
 Course Goals / Learning Outcomes
 Mode of Attendance & Participation Policy
 Grading & Grade Definitions
 Course Materials
 Academic Integrity Policy
 Accessibility Services Policy
 Publishing or Distributing Course Materials
 Policy

Case Work Product

Each case will have a description and specific instructions provided through the course [github repository](#).

Each student will work on case studies individually. Each will have the following work artifacts:

- Maximum 10min recorded slide presentation uploaded to youtube, embedded as a voiceover in the slides or in a similarly appropriate manner.
 1. The presentation will outline the business problem, the insights identified, describe the data and the analysis, and the outcomes/recommendations satisfying the course goals.
- Slide presentation uploaded to canvas (pptx file format)
- R Script(s) supporting the creation of any visuals, tables, and recommendations made during the presentation.
- Written supplemental describing problem, data, analysis, and recommendations.

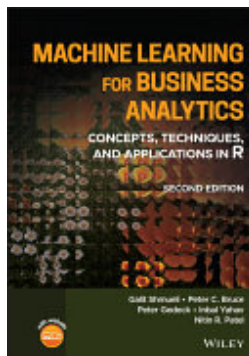
Essentially all supporting material including scripts, visual presentation slides, and/or written document will need to be submitted in for review.

The presentation will be evaluated on an equal-weighted basis with the following criteria.



- Organization – Was the presentation well organized?
- Delivery – Was the content delivered clearly and persuasively with the audience in mind?
- Code Documentation – Was the data mined to support the conclusion?
- Written Supplemental – Is the information clear and supported in narration and code? Did the information solve the case problem?
- Data Mining & Modeling Process – Overall, as a portfolio of work, is the topic interesting, organized, researched, supported and delivered effectively? CRISP-DM, SEMMA, or a similar workflow followed to organize the work?

Course Materials



Machine Learning for Business Analytics

ISBN: 9781119835172

Authors: Galit Shmueli, Peter Gedeck, Inbal Yahav, Nir Ailon
 MACHINE LEARNING FOR BUSINESS ANALYTICS Machine learning is a fundamental part of data science used by organizations in a wide range of arenas to turn raw data into actionable information. Machine Learning

Business Analytics: Concepts, Techniques, and Applications provides a comprehensive introduction and an overview of this methodology. This best-selling textbook covers both statistical and machine learning algorithms for prediction, classification, visualization, dimension reduction, clustering, recommendation mining, learning, experimentation, and reinforcement learning. Along with hands-on and real-life case studies, it also discusses managerial and ethical considerations for responsible use of machine learning techniques. This is the second edition of Machine Learning for Business Analytics. This edition also includes a new co-author, Peter Gedeck, with over 20 years of experience in machine learning using R. An expanded chapter is focused on discussion of deep learning techniques. A new chapter on experimental feedback techniques including A/B testing, uplift modeling, and reinforcement learning. A new chapter on responsible data science. Updated material based on feedback from instructors teaching MBA, Master of Business Analytics and related undergraduate, diploma and certificate courses, and from their students. A chapter devoted to relevant case studies with more than a dozen cases demonstrating applications for machine learning techniques. Each chapter exercises that help reinforce

gauge and expand their comp
and competency of the materi
presented A companion webs
more than two dozen data sets
instructor materials including e
solutions, slides, and case soli
textbook is an ideal resource f
level undergraduate and grad
Courses in data science, predi
analytics, and business analyt
also an excellent reference
researchers, and data science
practitioners working with qual
data in management, finance,
operations management, inform
systems, computer science, a
information technology.

Publisher: John Wiley & Sons

Publication Date: 2023-03-08

This textbook has some overla
lessons and should be purcha
students wishing to expand be
lessons to add additional fluer
technical knowledge. Each we
suggested readings and exerc
this textbook to reinforce and e
topics covered in class.

Academic Integrity Policy

You are responsible for understanding Harvard Extension School policies on [Academic Integrity](#) and how to use sources responsibly. Violations of academic integrity are taken very seriously. Visit [Using Sources Effectively and Responsibly](#) and [Harvard Guide to Using Sources](#) to review important information on academic citation rules.

Writing Code. While it may be common practice in many academic settings to adapt code examples found in textbooks, this is not the case in academia. In particular, you never copy code produced as coursework by other students, whether in the current term or a previous term, nor may you provide work for other students to use. Copying code from another student or any other source is a form of academic dishonesty, as is deriving a program substantially from that of another.

Writing code is similar to academic writing in that when you adapt code developed by someone else as part of your assigned coursework, you must cite your source. Plagiarism without proper citation is just as dishonest with programming as it is with prose. A program can be considered plagiarized even though no single line is identical to any line of the source code.

Accessibility Services Policy

The Division of Continuing Education (DCE) is committed to providing an accessible academic community. The [Accessibility Services Office \(ASO\)](#) is responsible for providing accommodations to students with disabilities. Students must request accommodations or adjustments through the ASO. Instructors cannot grant accommodation requests without ASO approval. It is imperative to be in touch with the ASO as early as possible to avoid delays in the provision of accommodations.

DCE takes student privacy seriously. Any medical document should be provided directly to the ASO if a substantial accommodation is required. If you miss class due to a sickness, notify your instructor and/or TA but do not include a doctor's note. Course staff will not request, accept, or review doctor's notes or other medical documentation. For more information, email accessibility@extension.harvard.edu.

[Course Information](#)
[Instructor Information & Office Hours](#)

Publishing or Distributing Course Materials

[Mode of Attendance & Participation Policy](#)
[Students may not post, publish, sell, or otherwise publicly distribute course materials without the written permission of the course instructor. Such materials include, but are not limited to, the following: lecture notes, lecture slides, video or audio recordings, assignments, problem sets, examinations, other student work, and answer keys. Students who sell, post, publish, or distribute course materials without written permission, whether for purposes of soliciting answers or otherwise, may be subject to disciplinary action, up to and including requirement to withdraw from the course. Further, students may not make video or audio recordings of class sessions for their own use without written permission of the instructor.](#)

Class Meeting Schedule

Please note that suggested tasks and assignments may be expanded upon within the class repository.

https://github.com/kwartler/Harvard_DataMining_Business

Adjustments will be made to the lessons based on the pace and priorities of the class.

September 4: NO CLASS (University holiday)

September 11: Class 1 - Introduction & Administrative
Introduction to R



September 18: Class 2 - Introduction to Data Mining, E

Assignments: Forum Introduction post (assuming we c
forum to Canvas) & Finish Skills Assessment

Suggested Reading: Chapters 1, 2

September 25: Class 3 More R Practice: Visualization
EDA

Assignments DUE (undergraduate only): HW1

Intro To R Homework.R Class Meeting Schedule

Suggested Reading: Chapter 3

October 2: Class 4 Data Mining in a business workflow
preprocessing, Donor Bureau Case

Suggested Reading: Chapter 6

October 9: NO CLASS (University holiday)

October 16: Class 5 Regression & Logistic Regression

Assignments DUE: Case I EDA Case

Suggested Reading: Chapters 6 & 10

October 23: Class 6 Decision Tree & Random Forest

Assignments DUE (mandatory for undergraduate & ext
graduate students): Visualization R Script

Suggested Reading: Chapter 9



Course Information

Instructor Information & Office Hours

October 30: Class 7 Time Series & Equity Trading

Course Goals / Learning Outcomes

Suggested Reading: Chapters 17, 18, 19

Mode of Attendance & Participation Policy

Grading & Grade Definitions

Course Materials

Academic Integrity Policy

November 6: Class 8 Consumer Credit, Risk Modeling
traditional Investment Modeling

Accessibility Services Policy

Publishing or Distributing Course Materials

Policy

Class Meeting Schedule

November 13: Class 9 Natural Language Processing (NLP)
of Words**November 20:** Class 10 Natural Language Processing (NLP)
How does chatGPT work & Document Classification

Assignment DUE: Case II Banking Case

November 27: Class 11 Possible Prerecorded Session
Data for your own projects: APIs & Webscraping**December 4:** Class 12 Unsupervised Clustering Analysis
Discriminant Analysis

Suggested Reading: Chapters 12, 16

December 11: Class 13 Hearing from industry professional in data space

Guest Speakers, awaiting confirmation



- Ross Leav, Presidio Ventures
- Rachel Switchenko, VP Customer Care Plymouth Assurance
- James Liu, Product Manager Amazon Web Services



December 18: (Final Exam or final class meeting) - Class session for final Case help