Course Syllabus: Data Mining for Business

CSCI E-96 – *although this course includes business topics at its heart it is a computer science course. Please be mindful of this.*

Harvard Extension Spring 2020
Dates: January 27, 2020 – May 16, 2020
Time: Monday 8-10pm
Building: Harvard Hall 101
Instructor: Ted Kwartler, MBA
Email:
edwardkwartler@fas.harvard.edu
Office Hrs: Available upon request
Optional Lab: Fridays 10AM EST (most weeks); ZOOM Link to be shared on Piazza

Important URLs:

Piazza (class discussion board, post LOTS of questions!)
https://piazza.com/class/k5co6lqzd9h53p

Piazza facilitates real classroom interaction although it is online. Piazza is used to realize distance learning, but it is not intended to be a customer service center, social networking website or the channel for sharing evaluations of teammates, staff or topics (save that for your course evaluations). Keep in mind teaching staff responses are not expected within any given timeframe. In fact, your teaching staff may not respond at all for some topics so that students have to help each other, research and explore on their own and ultimately learn rather than be explicitly told.

Canvas (homework submissions and grading)
https://canvas.harvard.edu/courses/69731

During class we will be trying out "helix" which uses zoom for video conferencing. Should we have any issues this is a back-up for chat questions:
https://canvas.harvard.edu/courses/69731/external_tools/1

Github repository allows you to get all scripts, powerpoints and data sets throughout the semester. For those not familiar with github, think of it like a shared drive similar to sharepoint or dropbox but with added functionality for data and computer science.
https://github.com/kwartler/Harvard_DataMining_Business_Student

Streaming Information:

Harvard Hall 101 is a brand new "Helix" room. Lectures are streamed live with multiple microphones and cameras around the room. Students will be able to access the live stream or recorded lecture via a link on the Lecture Video page. Recordings are usually posted within 24-48hours after the lecture. The live lecture becomes active approximately minute before the start of class.

Prerequisites:

- Textbook: Data Mining for Business Analytics: Concepts, Techniques, and Applications in R

ISBN-10: 1118879368

Harvard Coop Bookstore link for the book:

https://tinyurl.com/300-W20-CSCI-E-96-1

- Software: R & R-Studio
    a. If you are not familiar with R Studio please take a short introduction to R course at Lynda.com, DataQuest.com or DataCamp.com
- Access to git software to download data sets and class material or ability to download directly from the Internet
- A webcam or other method to record case presentations & upload to the University's approved site as well as interact with live class sessions if that is your mode of attendance
- Be prepared to obtain a free zoom account as each group will need a single zoom participant to record case presentations
- This semester we will be using https://rstudio.cloud/ to avoid local laptop issues for students. This will ensure all students are on the same environment and time won't be spent with technical troubleshooting. As a result, please sign up for a free account.

Course Learning Objectives:

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

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

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

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

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

Attendance:

Regular attendance and remote participation on the class forum is essential to the successful completion of this course. Attendance will be taken regularly for on campus sessions and forum participation will be monitored for remote participants. You are responsible for material covered in class even if you have not attended class or watched the recorded lectures. Given the amount of information covered, missing more than 1 class session for any reason may result in an automatic reduction in course grade. Unsatisfactory attendance may result in a failing grade. You should plan on spending at least three hours of independent study for each hour of class attendance.

Code of conduct:

This course expects you to uphold and report violations of the Extension School code of conduct found here. Further, all assignments are the responsibility of each *individual* pupil unless assigned as a group assignment. Utilizing the class forum, online resources, teaching assistants, and the class professor to ask questions is (of course) acceptable but copying another peer's work is considered a violation of the University code of conduct.

You are responsible for understanding Harvard Extension School policies on academic integrity (www.extension.harvard.edu/resources-policies/student-conduct/academic-integrity) and how to use sources responsibly. Not knowing the rules, misunderstanding the rules, running out of time, submitting "the wrong draft", or being overwhelmed with multiple demands are not acceptable excuses. There are no excuses for failure to uphold academic integrity. To support your learning about academic citation rules, please visit the Harvard Extension School Tips to Avoid Plagiarism (www.extension.harvard.edu/resources-policies/resources/tips-avoid-plagiarism), where you'll find links to the Harvard Guide to Using Sources and two, free, online 15-minute tutorials to test your knowledge of academic citation policy. The tutorials are anonymous open-learning tools.

Accessibility

The Extension School is committed to providing an accessible academic community. The Disability Services Office offers a variety of accommodations and services to students with documented disabilities. Please visit www.extension.harvard.edu/resources-policies/resources/disability-services-accessibility for more information.

Grading:

A course grade will be assigned on the basis of student performance on examinations, homework assignments, a written assignment, attendance and participation and group work. Students will

take their final exam online proctoring through a webcam. More details will be shared during class.

Homework is accepted up to 12 hours late. Any homework submitted after the deadline but before 12 additional hours will be penalized 10%. After 12 hours no late homework will be accepted under ANY circumstances. Pupils are expected to manage their own time and submit their work accordingly. Failure to submit submissions through the University approved portal by the assignment deadline will be considered late and not accepted. Submissions to any other location will not be accepted. During exams, no phones, tablets or computers should be used even as calculators. If you need a calculator you must bring one to your examination period. A student may prepare a single, double sided 3inch by 5inch, *handwritten* index card for use during any examination. Cards that are larger, typed or multiple cards will constitute cheating according to Harvard's academic integrity policies.

- Class participation, and online forum participation 10% of final grade. The course is a collaborative learning environment. The expectation is that *all* students will view, comment and ask questions on Piazza, request office hours as needed, and if offered attend or comment on Piazza for recorded lab sessions. Class participation is not free credit. If students do not contribute, they will not receive class participation credit.
- Case I 15% of final grade
- Case II 20% of final grade
- Final Exam 20% of final grade (Proctorio Window: Sun, May 10, 3 am to Tue, May 12, 3 am EST)
- Written assignment 15% of final grade
- Homework Assignments 20% of final grade. Homework will be due the Tuesday after it is assigned ie Tuesday (11:59 pm EST). Homework will be formatted according to this video concerning markdown and "knitr": https://piazza.com/class/jr99m5oqf5m6zi Poorly formatted homework is subject to a grading penalty of 10%.

Writing Assignment

Fifteen percent of the final grade will be determined by the quality and completeness of a 900 to 1200 word ***essay concerning ethical implications of data mining within a business context***. Approximately, no more than 25% of the essay should comprise a summary and synthesis of the assigned data science ethics articles. The balance of the essay can incorporate new literary sources and student reflections for how business is affected by the rise of cheap computing, large scale creation and storage of data and development of new algorithms. Example questions to spur creative reflection include (but are not limited to):

- Is it ok to have a "black box" algorithm where users do not know how it functions?
- Is there an ethical duty to tell users you are collecting information and reselling it or simply bury it in a terms of service agreement? Does anyone really read the agreements?

- Are algorithmic traders crowding out less sophisticated retail investors? Does the market have a duty to train others, disclose code based on open source licenses or report market manipulation?

While defining an ethical framework can be a personal matter, the organization and robustness of your argument along with supporting statements to the argument are subject to evaluation. It is not the case that all ethical actions are relative or that ethical considerations are incapable of objective evaluation. Further the level of sophistication you demonstrate in understanding the issue discussed, addressing applicable opposing viewpoints, actions stakeholders can take to mitigate issues and the logical structure of your essay will impact your grade. Lastly, primary source philosophical paradigms, not mere opinions should be used as a foundation for your logical construction of what is ethical in a data mining and business context.

Each page should have a header with a clear label including the author, date, page number and title. As a personal reflection paper concerning ethics, APA or similar citation method is *not* necessary.

Group Case Presentations

Students are expected to form groups of 3 or 4 by the end of week two to work on case studies. The teaching staff reserves the right to add new members to groups having only three or less members or to remove members from groups having five or more members. Each group will work on two business cases that use data to affect the outcome. Each group will create and upload verbal presentations for review and grading. During the recorded presentation, each individual in a group is expected to present a portion of the group's effort. Presentations will be graded on their use of data, code demonstration (if applicable), strategic business thinking, succinctness, persuasiveness, qualitative understanding of the business objective, and overall presentation skills. Each group presentation is to be no more than 15minutes in length. All supporting material including scripts, visuals and or presentation slides will also need to be turned in for review.

Once you join a case group you are not permitted to leave and join another. You are not permitted to work by yourself in a group of one. If you need help getting into a group, reach out to a TF.

Complaining about poor performing group peers is not helpful nor does it change the case expectations. In many real world scenarios, some teammates are stronger than others, leave jobs or have more time for a project. As a result, no adjustments to case evaluations are allowed no matter the circumstances. Graduate and adult course work at an elite college means you are expected to overcome any case difficulties. In the end, each team member is expected to perform an equal proportion of the case and the team is graded with this understanding.

Classes

Date	8-9pm	9-10pm	Reading Due	Assignments Due by Tuesday end of day (11:59 pm EST)
Jan 27	Introduction & Administrative	Intro to R	NA	
Feb 3	Intro to Data Mining	EDA	Chapter 1 Chapter 2	**HW 1** 1. Initial Reflection essay (12-15 sentences) 2. Piazza introduction post 3. C2.1 Data Mining Techniques 4. C2.2 Data Partition 5. C2.3 Data Sample 6. C2.4 Modeling Steps
Feb 10	Data Mining in a Business Workflow	Data Preprocessing Donor Bureau Case	Chapter 3	**HW 2** 7. Intro_To_R_Homework.R
Feb 17	**PRESIDENT'S DAY NO CLASS –** *work on your Case I*			
Feb 24	Regression		Chapter 6	**CASE I. OK Cupid Case** Upload
Mar 2	Logistic Regression	KNN	Chapter 7 Chapter 10	**HW 3** 8. C6.1 Predicting Boston Housing Prices Only do a, & b
Mar 9	Decision Tree	Random Forest	Chapter 9	**HW 4** 9. C10.3 Sales of Riding Mowers Only do a, b,c & d 10.C9.4 Personal Loan Acceptance
Mar 16	**Spring Break**			
Mar 23	Time Series Forecasting		Chapter 16, 17, & 18	**HW 5** 11. C9.3 Predicting Prices of Used Cars Only do "a"
Mar 30	Equity Trading & Portfolio Optimization (time permitting)			**HW 6** 12. C16.1 Impact of 9/11 on Air Travel *use Sept11Travel_REVISED.csv 13. C18.9 Australia Wine Sales Only do "a", & "b"
Apr 6	Financial Risk Modeling; following a full SEMMA & Champion/Challenger Workflow	Non-Traditional Investment Modeling	NA	**HW 7** 14. 1_REVISED_TTR_homework.R for any equity not covered in class.*
Apr 13	Intro to Natural Language Processing			Nothing to submit this week - Work on your Case II
Apr 20	Data Sources with R - APIs	Reporting Automation	NA	**HW 8** 15. Using text data provided, build a

		w/Powerpoint & Flexdashboard		wordcloud, comparison.cloud and commonality cloud 16. C18.9 Australia Wine Sales Only do "a", & "b"
Apr 27	Collaborative Filtering	Ethics	Chapter 14	**HW 9** 17. Create a script to construct a powerpoint with lib(officer)* 18. Create a script to construct a flexdashboard* * *more details shared in class*
May 4	Guest Speakers, *awaiting confirmation* • Ross Leav Presidio Capital • Angela Chow Analytics Manager Wayfair • Gaurav Rao, Head of Product, Neural Magic			**CASE II. Banking Case Upload**
May 7 (Thurs)	NA- not a class session			**HW 10** 19. Final reflection essay (12-15 sentences) 20. C14.2 Identifying Course Combinations
May 10 (Sun)	**Final Exam** Remote proctored exam will be available from Sun, May 10, 3 am to Tue, May 12, 3 am EDT **48-hour** time window (not 72-hour window)			**Final Exam**
May 14 (Thurs)	NA – not a class session			**Writing Assignment**

Graduate Credit Students

This course is open to non-credit, graduate and undergraduate students. As a result, the course experience will vary for each cohort.

Noncredit students may submit case presentations, homework, and the ethics paper. Your assignments will receive feedback to improve your acumen. However noncredit student may not take exams or receive letter grades.

Graduate credit students are expected to do more work and perform at higher standards than undergraduate credit students. On exams, there will be additional knowledge tested for graduate credit students. These may include but are not limited to additional multiple choice questions, short form answers or coding sections. Further, a graduate credit student's ethics paper should incorporate an additional 3 sources of information beyond the assigned reading. Similarly a graduate writing paper needs to demonstrate nuanced sophistication of the ethical considerations presented along with appropriate counter arguments

Grading Scale

You earn the grade based on assignments according to the scale below. Grades are not curved to fit a predetermined distribution. A student's degree, certificate candidacy, or funding status will

not have any impact on a course grade. "Needing an A" for any reason is not sufficient to earn an A grade. *Note there are no "minus" grades given in the course.* It is the belief of the instructor that minus grades constitute a false precision in many academic courses and further penalize frequent "A-" students since there is no way to obtain an "A+" to rebalance a GPA. To the student's benefit, one can still earn a "plus" on their final grade according to 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

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder