UNIVERSITY OF CALIFORNIA, DAVIS

GRADUATE SCHOOL OF MANAGEMENT

MACHINE LEARNING

(BAX-452, WINTER 2024)

PRELIMINARY SYLLABUS (12/14/2023)

Instructor: Jörn Boehnke, Assistant Professor

Office: 3404, Gallagher Hall, Graduate School of Management

E-Mail: jb@ucdavis.edu

Office Hours: Every Monday 7:15am – 8:15am at
https://ucdavis.zoom.us/j/96665507050?pwd=ckhydkJPYmw0UlZFSmtCbUNVd2FsZz09

The first office hour is 1/8.

Please feel free to reach out to me via email at any point. We can also schedule additional meetings if necessary.

Objectives: After taking this course, you will know how to explore and analyze large high-dimensional datasets, become adept at building powerful systems for prediction, and gain the understanding necessary for interpreting the structure in such models. The topics include:

Introduction
Regression
Treatment Effect
Bootstrap
Model Selection
Classification
Clustering
Factors
Trees
Neural Networks
Natural Language Processing
Large Language Models
Reinforcement Learning*
Spatiotemporal Analysis*

* topic may be shortened or skipped in the interest of time

We learn both basic underlying concepts and practical computational skills, including techniques for analysis of distributed data. Strong emphasis is placed on the analysis of actual datasets, and on the development of application specific methodology. Among other examples, we will consider consumer data mining, internet and social media tracking, sports analytics, and text mining.

TAs: Ritesh Chauhan (ritchauhan@ucdavis.edu)
Kuang Li (kuali@ucdavis.edu)
Tiruo Yan (tyyan@ucdavis.edu)

The TAs offer weekly office hours and occasional seminars. Details are TBD.

Please also feel free to reach out to your TAs via email at any point.

Evaluation:
15% Assignments
5% Mini exam week 1 (Monday, 1/8 5pm)
25% Exam week 6 (Monday, 2/12 5pm)
25% Exam week 10 (Monday, 3/11 5pm)
15% Final group project
15% Class participation

Late assignments, exams, and projects will not be accepted.

Clerical scoring errors will be corrected without hassle, but for other re-grades you must hand back the work and submit an email request; the entire paper will be subject to re-grading.

Groups: You will be assigned to a group at the beginning of the quarter to discuss assignments and work on the final project. *Please note that group work will be subject to a peer assessment at the end. One-half of your total group work grade (assignments and final project) will be determined by the score on the peer evaluations.* I.e., individuals who do not contribute their fair share to the group (as determined by the group) will be penalized.

Class Participation: I hope you actively participate in class. Meaningful class interactions add greatly to the learning experience, and I consider class participation to be an important component of this course and its success. Accordingly, active and meaningful participation in class and on Piazza is part of your grade evaluation. Please be prepared to be cold-called in class.

Assignments: There will be 8 one-week homework assignments. Homework assignments are released after class. *Only the best 7 assignments will be counted towards your grade.* Assignments are submitted through Canvas and should have a clear and concise

presentation.

All assignments are individual assignments. Assignments *may be* discussed among your group members but *must be* submitted individually. While your code submission may be similar to that of your group's members, any write-up, code comments, etc. must be your own. Simply copying and pasting the write-up and code comments from other members in your group will yield point deductions. I want each group member to fully understand the code and analytics required in assignments.

Assignments are not to be discussed outside your group. Any violation of this policy will be considered plagiarism and penalized as such.

Classroom Etiquette / Strikes / Extra Assignments: Classes begin on time. You are required to attend every class unless extenuating circumstances arise and make it difficult to do so. If you are unable to attend a class, you are required to inform me about the extenuating situation by email *at least one hour* before the start of class. A late arrival ($\geqslant$ 5 minutes) or skipped class results in *one strike*.

All cell phones must be muted before the start of class. The computer should only be used to take class notes and code / run in-class exercises. All other programs should be shut down before the start of class. Any messaging, web surfing, e-mailing, etc. disturbs the class and is a breach of classroom etiquette. If it comes to my attention, any such behavior results in *one strike*.

Strikes are accumulated throughout this course. Ideally, your strike balance remains zero. However, I recognize that the unexpected sometimes happens, and a total of two strikes or less will remain without consequences. Any strike above two yields an additional individual assignment. I.e., the number of (counted) assignment goes up by one for each strike in excess of two.

Exams: There will be three exams. The exams will be conducted on Monday 1/8, 2/12, and 3/11 at 5pm Pacific Time using Respondus LockDown Browser + Monitor. All exams will be closed book exams.

Exam 1 will cover the core concepts learned in Fall Quarter's BAX-401 and BAX-441. Exams 2 and 3 will cover the concepts learned in this course.

Final: In the final project you will study a business problem of your choice. The goal of the project is to describe and analyze data for business use. Specifically, your analysis ought to add insights that help the business problem at hand. For this project, think of yourselves as consultants who want to support some company's business decisions. These decisions should be based on insights gained from analyzing the data and it is your task to provide this analysis.

Planning should begin around week 6. Successful projects will require statistical programming and managerial skills. They will implement a variety of machine learning

concepts we covered in class. All implementations must be documented, and all code must be attached to your submission. Moreover, this project must be significantly different to any final project developed for other MSBA classes.

Any number of primary and secondary data sources can be used for this project. You are allowed *but not required* to use the data from the final project of BAX-422: Data Design and Representation as the data source for this project. The project reports must be fully independent for the two courses in any case. Please always adhere to the terms and conditions pertaining to the websites and data you use.

Typically, project reports will involve the following components: (1) Title Page, (2) Executive Summary, (3) Background, Context, and Domain Knowledge: business (scenario) in mind, type of industry, products / services, competition, (4) Discussion of how the firm / industry traditionally attempts to solve this problem, the specific strategy, and how it aligns with the business model, (5.1) Analyses, (5.2) More Analyses, (5.3) Even More Analyses, (6) Recommendations and Business Value provided, and finally, (7) Summary and Conclusions.
(5) and (6) will showcase what you learned in class and how well you can apply it. Please make sure that these components form the centerpiece of your project report.

The final project report cannot exceed 10 pages (not including the title page, code attachments, references, and appendices). Any necessary tables, figures, visualizations, and text must be contained within the 10 pages. I urge you to ensure that the written report is direct, insightful, and specific to the problem at hand. The report should adhere by the following formatting guidelines: text to no smaller than 11-point font, 1-inch margins on all pages, and all text should be double-spaced. The report should also contain an executive summary, which counts towards the page limit. **All submissions are due on Thursday 11:59pm Pacific of finals week (i.e., Thursday of week 11).** No exceptions.

Software: In the evolving landscape of business analytics, Python has emerged as the predominant programming language, widely recognized for its versatility and effectiveness. Reflecting this industry trend, our course will now exclusively use Python for all computing tasks (for comparison: it was 50% Python and 50% R last year). This focus will allow us to dive deeply into its application in business analytics, providing you with relevant and practical skills.

Python's extensive use in data analysis and its robust libraries make it a great tool for business analysts. During this course, all homework and class examples will be in Python, ensuring a consistent and focused learning experience. While a basic understanding of Python is expected at the start of the class, be prepared for a steep learning curve if you are new to it.

This course is not an introduction to Python but an application of Python in business

analytics. I encourage you to install Python and familiarize yourself with it as soon as possible. Our goal is to equip you with the Python skills that are highly sought after in today's business environment.

Rules: All participants in the course, instructor and students, are expected to follow the UC Davis Principles of Community, which includes affirmation of the right of freedom of expression, and rejection of discrimination. The right to express points-of-view without fear of retaliation or censorship is a cornerstone of academic freedom. A diversity of opinions with respectful disagreement and informed debate enriches learning. However, in this course, any expression or disagreement should adhere to the obligations we have toward each other to build and maintain a climate of mutual respect and caring.

You are expected to take UC Davis's Code of Academic Conduct as seriously as we do. You were given this code of conduct with explicit explanations of violations (e.g. plagiarism, cheating, unauthorized collaboration, etc.) and your responsibilities in regard to them during orientation, and you signed a statement affirming that you understand it. Academic conduct violations will not be tolerated, and your instructor will not hesitate to turn violators over to Student Judicial Affairs. If you are uncertain about what constitutes an academic conduct violation, please refer to the code linked above, contact your instructor, or refer to the Office of Student Judicial Affairs.

Q&A: We are using Piazza for class discussion (accessible through Canvas). The system is highly catered to getting you help fast and efficiently from classmates, the TAs, and myself. Rather than emailing your questions, try to post on Piazza. If you have any problems, you can email me or team@piazza.com.

Please feel encouraged to answer your classmate's questions; it's a huge help to us, and even if you are wrong everyone learns (we check the answers and clear up confusion). While you can post anonymously, we encourage you to take credit for your questions and answers.

Texts: There is no required textbook; all required materials will be available on the class page. The best preparation you can do before lectures is to go through class code and work through examples.

A good introductory book is *An Introduction to Statistical Learning: with Applications in Python*, by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor.

A great advanced text is *The Elements of Statistical Learning* by Trevor Hastie, Robert Tibshirani, and Jerome Friedman, but it requires some mathematical sophistication and goes beyond the material we will be covering. Also good, but still advanced, are *Pattern Recognition & Machine Learning* by Christopher M. Bishop and (the encyclopedic) *Machine Learning* by Kevin P. Murphy. Both books introduce the material from a more computer engineering (or AI), rather than statistical, perspective.

Accommodation: UC Davis is committed to educational equity in the academic setting, and in serving a diverse student body. All students who are interested in learning about how disabilities are accommodated can visit the Student Disability Center (SDC). If you are a student who requires academic accommodations, please contact the SDC directly at sdc@ucdavis.edu or 530-752-3184. If you receive an SDC Letter of Accommodation, submit it to your instructor for each course as soon as possible, at least within the first two weeks of a course.

Safety and Emergency Preparedness: UC Davis has many resources to help in case of emergency or crisis. While reviewing campus Emergency Information, you may want to register for UC Davis Warn Me and Aggie Alert, which will give you timely information and instructions about emergencies and situations on campus that affect your safety.

If there is an emergency in the classroom or in non-Davis locations, follow the instructions of your instructor.

Disclaimer: Unexpected events might require elements of this syllabus to change. Your instructor will keep you informed of any changes.