EDUC 6191: Core Methods in Educational Data Mining
Fall 2024
Instructors: Ryan Baker and Joyce Zhang

SYLLABUS

Instructor Info
Email: rybaker@upenn.edu , joycez@upenn.edu
Virtual office hours for Ryan: Thursdays 3p-4p
https://upenn.zoom.us/j/91206129073
Virtual office hours for Joyce: Mondays 6p-7p
https://upenn.zoom.us/j/97788186360
Course time: Thursdays 615p-805p
https://upenn.zoom.us/j/93366352388
Class discussion forum: https://educ6191-f24.jeepytee.net/

Required Texts:

- Baker, R.S. (2024) *Big Data and Education,* 8th edition
https://www.upenn.edu/learninganalytics/MOOT/bigdataeducation.html

Information on how to obtain other course readings will be provided in class. Most are available on the course schedule webpage.

Course Goals: This course covers core methods in educational data mining. Students will learn how to execute these methods in standard software packages, and the limitations of existing implementations of these methods. Equally importantly, students will learn when and why to use these methods. Discussion of how EDM differs from more traditional statistical and psychometric approaches will be a key part of this course; in particular, we will study how many of the same statistical and mathematical approaches are used in different ways in these research communities.

Course Pre-requisites: Python, EDUC7667 or instructor permission. You must be able to use Python before taking this course. If you are not able to use Python, do not take this course. Professor Bodong Chen will be offering a more introductory course that does not require Python. If you are in the Masters in Learning Analytics, Professor Chen's course does not substitute for this course; please contact Professor Baker and Dr. Li.

Assignments: No exams or quizzes will be given.

Across the semester, eight basic assignments and four creative assignments will be assigned. The assignments are listed below in the class schedule. Every student must do six of the basic homework assignments (your choice) and three of the creative homework assignments and turn them in by posting them to the discussion forum. For the basic homework assignments, you must complete 3 of assignments 1-4, and 3 of assignments 5-8. You may choose any three creative homework assignments to complete.

If you do more assignments than required, I will give you feedback, but I will not grade extra assignments and let you have the ones with a higher grade. If you turn in all 4 of the first 4 basic assignments, I will only grade 1-3. If you turn in all 4 of the second 4 basic assignments, I will only grade

5-7. If you turn in all 4 of the creative assignments, I will only grade 1,2, and 3. To repeat, you cannot get a higher grade in this class by doing more assignments. You cannot get extra credit by doing more assignments. The point of this policy is to let you choose which material to focus your energy on, and to do a better job on that material.

Participation in asynchronous activities will also be part of the course grade. For each creative assignment you complete, you are expected to also provide substantive comments on at least four other students' submissions. For these posts, there is no length requirement, but the posts must offer a critical and meaningful perspective on how that student did the assignment. To count towards your grade, your posts must be submitted within five days of the assignment being posted.

In addition, you are expected to watch the asynchronous videos and post questions and comments on each week's content, or ask questions and comments in the in-person class. You will not be graded on the content or quantity of this type of participation, but if you don't do this, you won't get nearly as much out of the class.

Extensions will be given on the assignments as needed. However, please be reasonable.

Grading

- 6 of 8 Basic Assignments 6% each (up to a maximum of 36%)
- 3 of 4 Creative Assignments 13% each (up to a maximum of 39%)
- Asynchronous participation 8% each (up to a maximum of 24%)
- 1 bonus point

BONUS: For every creative assignment, there will be a special bonus of 20% for the best hand-in. "Best" will be defined specifically in each assignment.

Foundation model policy: Within this class, you are welcome to use foundation models (ChatGPT, GPT, DALL-E, Stable Diffusion, Midjourney, GitHub Copilot, Claude, Meta, Mistral, and anything after) in a totally unrestricted fashion, for any purpose, at no penalty. However, you should note that all large language models still have a tendency to make up incorrect facts and fake citations, code generation models have a tendency to produce inaccurate outputs, and image generation models can occasionally come up with highly offensive products. You will be responsible for any inaccurate, biased, offensive, or otherwise unethical content you submit regardless of whether it originally comes from you or a foundation model. If you use a foundation model, its contribution must be acknowledged in the handin; you will be penalized for using a foundation model without acknowledgement. Having said all these disclaimers, the use of foundation models is encouraged, as it may make it possible for you to submit assignments with higher quality, in less time.

Plagiarism policy: The university's policy on plagiarism still applies to any uncited or improperly cited use of work by other human beings, or submission of work by other human beings as your own. If you are not sure whether some action counts as plagiarism, ask before doing it. The university's policy on plagiarism will be strictly followed.