COMPSCI 590ED: Educational Data Mining and Learner Analytics

I. COURSE CATALOG

This course provides an overview of the emerging field of educational data mining and data analytics over educational data sets. Some of these datasets will come from student users of learning technologies, and are a reflection of students abilities and other states of mind as they learn new material. We will discuss both the application of machine learning methods to educational data, as well as classic models inspired by cognitive theory, as well as exploring and finding patterns in data. The course will use a combination of lectures, paper presentations, lab assignments, and projects.

II. COURSE DESCRIPTION

The primary goal of this course is for students to be comfortable with exploring educational datasets and analyzing them, finding patterns in educational data, creating models that summarize and describe data, and creating predictive models, using a variety of techniques that span machine learning, data mining, and statistics. We will work with datasets that come from educational learning technologies, tutoring and assessment software, as well as other datasets that students might be interested in (e.g. datasets from students grades classes, many of which are publicly available through national sources). Students are encouraged to bring their own datasets of interest, if they have them or want to. Another goal is to understand how current researchers in the field of educational data mining, AI in Education, and educational technology are using a variety of techniques to understand and model their data. We will cover many different methods, from machine learning methods applied to educational data and classic models inspired by cognitive theory. We will also cover various applications including computerized adaptive testing, affect detection, reading and writing analysis, activity log analysis, and instructional policy design.

Required background: basic knowledge of programming, basic knowledge of probability and statistics.

III. COURSE DETAILS

Instructors: Ivon Arroyo and Andrew Lan
TA: William Rebelsky, wrebelsky@umass.edu
Class meeting time & location: TTh 1:00-2:15PM, CS 142
Office Hours: W 5-6PM, CS 230
Credits: 3
Course webpage (tentative): https://people.umass.edu/~andrewlan/courses/590EDfall21.html; course materials will be on Moodle
Prerequisites:
Undergraduates: COMPSCI 240 (or STATISTCS 515) with a grade of C or better
CS Graduates: Undergraduate preparation assumed
EDUC Graduates: EDUC 555 (or EDUC 656) with a grade of C or better
Textbooks (optional):
Hand, Mannila, and Smyth (2001) Principles of Data Mining, MIT Press.
Goodfellow, Bengio, Courville, and Bengio (2016) Deep Learning, MIT Press.
Romero, Ventura, Pechenizkiy, and Baker (2010) Handbook of Educational Data Mining.

IV. GRADING

- 4 total homeworks and programming assignments, 30%
- In-class presentations, 30%
- Midterm project plan, 10%
- Final project report, 30%
- **Tentative** grading scale: A: 90-100%, B: 80-89%, C: 70-79%, D: $\leq$ 70%

V. TENTATIVE SCHEDULE

The course will be in a hybrid format involving lecturing and paper presentations. The first part of the course will cover background material, involving more lectures. The second part of the course will cover applications, involving more paper presentations.

- Week 1: Introduction to EDM, intelligent tutoring systems, visualizations, statistics
- Exploratory Methods:
- Week 2: Clustering, HW1
- Week 3: Association Rule Mining
- Week 4: Linear and nonlinear regression, HW2
- Week 5: Bayesian Classifiers and Bayesian networks
- Week 6: Neural Networks
- Week 7: Midterm project presentations
- Week 8: Item Response Theory and Computer Adaptive Testing
- Week 9: Knowledge tracing, HW3
- Week 10: Learning content analysis
- Week 11: Activity log analysis, HW4
- Week 12: Instructional policy learning
- Week 13: Final project presentations

VI. DETAILS ON COURSE COMPONENTS

- Midterm: Individual. Posing one (or a few alternative) creative research question(s) to ask to a dataset of the students' choice, and possible techniques to answer it. Writeup of a discussion of why the question is interesting, and what data could be used to answer it, and what techniques could be used and how. 5 Pages.
- Final project: Individual. You will choose one of the research questions from the mid-term, and will apply Educational Data Mining Techniques to answer a research question(s) to ask to a set of data of students' choice, and being able to show the process to answer the question, using at least one technique. The person will should submit a single 10-15 page write-up (can build up from the midterm document). Final Project will also involve a presentation of the final project.
- Individual presentations: Individual students will be assigned research articles every couple of weeks, about twice a semester (depending on the number of students). Students should plan for a 20 minute presentation of the research article with discussion that could span 5-10 minutes. Sign-up sheet on Canvas. All students are required to read the research articles before class. Students should present discussion points at the end of their presentation.
- Class participation: In-class participation and in bulletin boards and online discussions. Students will read the research articles ahead of time, and think of questions/items to discuss in class.

VII. ACCOMMODATION STATEMENT

The University of Massachusetts Amherst is committed to providing an equal educational opportunity for all students. If you have a documented physical, psychological, or learning disability on file with Disability Services (DS), you may be eligible for reasonable academic accommodations to help you succeed in this course. If you have a documented disability that requires an accommodation, please notify me within the first two weeks of the semester so that we may make appropriate arrangements.

VIII. ACADEMIC HONESTY STATEMENT

Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst. Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent. See https://www.umass.edu/honesty/.