INFO 5200 Introduction to Learning Analytics | Spring 2019

Cornell Computing and Information Science Professor Rene Kizilcec

Mondays and Wednesdays, 8:40 – 9:55am in Snee Hall, room 1150

Introductory graduate-level course on learning analytics with a focus on learning science theories, data mining and prediction methods, and hands-on applications with fine-grained learner data.

Instructor

Professor Rene Kizilcec
Email: kizilcec@cornell.edu
Office: 208 Gates Hall
Office Hours: Mondays, 3-4pm, or by appointment

Teaching assistant: Upol Ehsan <ue28@cornell.edu>
TA Office Hours: TBD

Course Description

Technology is transforming the ways we learn and teach. Learning is increasingly occurring on the Internet or through educational software, which has led to an explosion of educational data. How can all of the interaction, discourse, and performance data that are generated from online courses, learning management systems, and student discussion forums be used to improve educational effectiveness and support basic research on learning? This introductory course on Learning Analytics provides a survey of educational data science methods (predictive modeling, classification, regression, causal inference) and learning science theories (active learning, modalities, Bloom's taxonomy, metacognition, self-regulated learning). Students will collect and analyze their own learning trace data as part of the course.

In this course, you will learn:

1. key insights from research in the learning sciences on how learning works
2. how to select and apply methods from educational data mining and learning analytics to analyze different kinds of educational data, and interpret its results
3. the strengths and weaknesses of these methods for different applications
4. the potential benefits and risks of learning analytics for students, teachers, and institutions

Who Should Take this Course

This course is designed for MPS students and undergraduate (honors) students who are interested in learning, educational technology, methods for analyzing educational data, and the broader implications of these technologies in education. The course will involve guided data analysis using the R statistical programming language. Prior knowledge is a key predictor of learning outcomes according to learning science research. Some knowledge of either statistics, data mining, mathematical modeling, or algorithms is recommended. Ideally (though this is not required), you have some experience working with data in R.

Course Structure, Workload, and Grading

The course meets twice a week for 1 hour 15 minutes. Punctual attendance is expected. Classes will comprise a variety of learning modalities such as interactive lectures putting topics from the readings into context and informed by student responses to the reading; case study presentation and class group discussion; and hands-on in-class problem solving for skill building.

There will be a **weekly reading** in the form of a chapter from the *Handbook of Learning Analytics* or an academic article to be completed *before class begins*. Students are asked to answer a few **comprehension questions** about the reading, which are graded in terms of correctness and count for 10% of the total grade. Students also submit a **two-paragraph reaction** online, which is graded in terms of completion (0/1) and counts for 10% of the total grade. In your reaction, you may (1) ask questions about points that are unclear in the reading, (2) raise issues that you think we should discuss, or (3) comment on ideas that you think are interesting.

In most weeks of the course, there will be a **homework assignment** that requires preparation or analysis of a dataset using the methods introduced in class that week. The dataset and starter code will be provided online. Responses should be uploaded online and count for 40% of the total grade.

There will be a two course projects (a group project and an individual project) that count together for 40% of the grade.

Course page, FAQs, readings, assessments, and discussion are available on EdX Edge:
https://edge.edx.org/courses/course-v1:CornellX+INFO5200+2019_Spring/about

Textbook / Readings

There will be assigned reading in preparation for each class. Most of the readings are from the Handbook of Learning Analytics (HLA) which is freely available online. Additional readings based on academic articles and books will be provided. There is no need to purchase the books below.

Lang, C., Siemens, G., Wise, A., & Gasevic, D. (2017). The Handbook of Learning Analytics – First Edition. SoLAR.

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. John Wiley & Sons.

Schwartz, D. L., Tsang, J. M., & Blair, K. P. (2016). *The ABCs of how we learn: 26 scientifically proven approaches, how they work, and when to use them*. WW Norton & Company.

Sclater, N. (2017). *Learning analytics explained*. Routledge.

Accommodations and Counseling Resources

Let me know as soon as possible if you require special accommodations due to learning disabilities, religious practices, physical requirements, medical needs, etc.

It is not uncommon for students to experience stressful events at some point during graduate school. Students sometimes experience depression, anxiety, family stress, the loss of loved ones, financial strain, and other stressors. It is perfectly normal for students to seek the service of mental health professionals to provide them with support and skills to cope with these experiences. Below I have provided the contact information for some of the mental health services available to Cornell University students so that you will know where you can go if you or a friend would like to take advantage of these resources.

Cornell Health
110 Ho Plaza, Ithaca, NY 14853-3101
Phone: (607) 255-5155

Academic Integrity

Academic integrity is a serious matter and will be treated as such. Students should be aware that violations of academic integrity can impede the development of trust within our community, and that penalties can be severe.

Each student in this course is expected to abide by the Cornell University Code of Academic Integrity at all times. Any work submitted by a student in this course for academic credit must be that student's own work. All outside assistance should be reported, and the work of others should be properly cited. Students should be aware that cultural differences may exist in standards and definitions regarding academic integrity. I apply the Cornell University standards in this class. Misrepresenting another's work as your own means submitting or presenting somebody else's words or ideas without proper attribution. Proper attribution includes quotation marks and page numbers for any words taken directly from any piece of another author's work, and/or a citation when you have paraphrased or

summarized somebody else's work. Sources need not be published to be cited; any document that you use as a source – even if you are the author - must be cited or attributed in this way.

Detailed Class Schedule

Week 1: Introduction

Jan 23, 2019: Course Overview

Topic

Motivating examples of why learning analytics is important; in-class activity; course outline, logistics, and systems; introducing first group project.

Homework *(due Friday 8pm)*

1. Access Edx Edge course site
2. Set up R and RStudio, run test script, answer questions, take R tutorial (if applicable)
3. Start work on group project about How Learning Works; read chapter, create slides, and practice for presentation in week 3

Week 2: Overview of Learning Analytics

Jan 28, 2019: Learning analytics and educational data mining
Jan 30, 2019: Data types and overview of methods

Topic

What is learning analytics? Why does it matter? Why should you care about it? Where is it all going? Why measure learning? Is it all about learning? What does data from educational systems look like? What is unique about educational data? What methods are commonly used to analyze data in education? What is the difference between descriptive and inferential analysis?

Reading *(before class)*

1. N Sclater, Learning Analytics Explained, Ch.1 (by Monday)
2. N Sclater, Learning Analytics Explained, Ch.8 (by Wednesday)
3. R S Baker & G Siemens, Educational Data Mining and Learning Analytics (by Wednesday)

Homework *(due Friday 8pm)*

1. Load a dataset into R and create a report with basic descriptive statistics using starter code posted online
2. Explore two different datasets posted and answer the same basic questions with each dataset (make plots, compute differences, compute correlations, engineer features)
3. Self-assess your progress

Week 3: How Learning Works

Feb 4, 2019: Group presentations and discussion on How Learning Works Pr. 1-4
Feb 6, 2019: Continue presentations and discussion on How Learning Works Pr. 5-7

Topic
What do we know about how learning works? Students present in groups about the seven research-based principles in How Learning Works.

Homework *(due Friday 8pm)*
1. Read the chapter of How Learning Works on the principle you find most interesting
2. Write about 2-3 concrete ways you could apply the principle in an introductory programming class. Explain how you implement it and why you think this implementation should work. How might you assess if it is actually working?
3. Submit your response on a Google slide and leave a constructive/encouraging comment on another student's slide

Week 4: Predictive Modeling
Feb 11, 2019: Overview of predictive modeling uses and approaches
Feb 13, 2019: Prediction in R

Topic
Predicting learner behavior and learning outcomes is a cornerstone of learning analytics. There are various methods to choose from depending on the prediction task and available data.
What teaching/learning techniques can support learning? How can technology support learning and what are examples of such technologies?

Reading *(before class)*
Handbook of Learning Analytics, Ch.5, Predictive Modelling in Teaching and Learning

Homework *(due Friday 8pm)*
Compare the performance of several predictive models on two datasets

Week 5: Assessment
Feb 18, 2019: Guest speaker: Dr. Julie Miles, Houghton Mifflin Harcourt (hmhco.com)
Feb 20, 2019: Modeling learning with Bayesian knowledge tracing

Topic
What exactly are is the data produced by learning environments measuring? How to measure learning? Developing robust assessments. Knowledge tracing.

Reading *(before class)*
1. Handbook of Learning Analytics, Ch.1, Theory and Learning Analytics *(by Monday)*

2. Handbook of Learning Analytics, Ch.3, Measurement and its uses in learning analytics *(by Monday)*
3. Corbett & Anderson, 1994, "Knowledge tracing: Modeling the acquisition of procedural knowledge" *(by Wednesday)*

Homework *(by Friday 8pm)*
Create a new knowledge assessment, collect data, and evaluate the psychometric properties of the assessment

Week 6: Predictive Modeling cont.
Feb 25, 2019 – February Break
Feb 27, 2019: Deep knowledge tracing, and class project

Reading *(before class)*
Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In Advances in Neural Information Processing Systems (pp. 505-513).

Week 7: Predictive Modeling Project
Mar 4, 2019 – No class, see online materials (LAK conference)
Mar 6, 2019 – No class, see online materials (LAK conference)

Topic
Predictive analytics class project

Homework *(due Friday 8pm)*
Complete project and submit project report

Week 8: Causal Inference
Mar 11, 2019: Why experiments and how they work?
Mar 13, 2019: Design and analysis of experiments

Topic
Randomized experiments are the most effective way to estimate the effects of educational interventions. We will review why they work, how to design them well, and analyze the results.

Reading *(before class)*
1. J Pearl, The Book of Why, Ch.1
2. Handbook of Learning Analytics, Ch.18, Diverse Big Data and Randomized Field Experiments in MOOCs

Homework *(due Friday 8pm)*

1. Design an experiment to test a learning science principle
2. Analyze data from an experiment

Week 9: Self-regulated Learning
Mar 18, 2019: Self-regulated learning (SRL)
Mar 20, 2019: Measuring and supporting SRL strategies

Topic
The self-regulated learning (SRL) literature describes

Reading *(before class)*
Handbook of Learning Analytics, Ch.21, Learning Analytics for Self-Regulated Learning

Homework *(due Friday 8pm)*
1. Keep a diary of your own SRL activities
2. Analyze a dataset to identify evidence of self-regulation

Week 10: Multimedia Learning and Video Analytics
Mar 25, 2019: Introduction to multimedia learning
Mar 27, 2019: No class, see online materials (Cornell Silicon Valley event)

Topic
Video has become a popular medium in online learning and the data from how learners interact with videos can yield actionable insight to better support learners.

Reading *(before class)*
1. R C Clark & R E Mayer, e-Learning and the Science of Instruction, Ch.2 and 4
2. Handbook of Learning Analytics, Ch.22, Analytics of video use

Homework *(due Friday 8pm)*
Analyze video analytics data

Spring Break
April 1-5, 2019

Week 11: Dashboards
Apr 8, 2019: What makes a good dashboard?
Apr 10, 2019: In-class dashboard design activity

Topic
Dashboards are a way to present analytics to stakeholders (students, teachers, administrators) to inform their actions.

Reading *(before class)*
Handbook of Learning Analytics, Ch.12, Learning analytics dashboards

Homework *(due Friday 8pm)*
Make a dashboard with R shiny

Week 12: College Pathway Analytics
Apr 15, 2019: Analytics for higher education
Apr 17, 2019: Sequences analysis to understand pathways

Topic
College students make choices about which courses to choose. What information informs these choices and how can macro-level learning analytics in higher education shape college pathways.

Reading *(before class)*
1. Stevens et al. (2018), "Choices, Identities, Paths: Understanding College Students' Academic Decisions"
2. Chaturapruek et al. (2018), "How a data-driven course planning tool affects college students' GPA", ACM Learning at Scale

Homework *(due Friday 8pm)*
Analyze course sequence data to visualize pathways and/or identify toxic course pairings

Week 13: Emotional learning analytics
Apr 22, 2019: Emotion and affect in learning
Apr 24, 2019: Predicting learner affect

Topic
Emotional learning analytics; what is affect, why does it matter, how can it be used?

Reading *(before class)*
Handbook of Learning Analytics, Ch.10, Emotional learning analytics

Homework *(due Friday 8pm)*
Train an affect detector using dataset provided

Week 14: Ethics and Looking Ahead
Apr 29, 2019: Ethical considerations of learning analytics
May 1, 2019: Where learning analytics is going and course wrap-up

Topic

What are the limits of learning analytics and where should limits be imposed? Data and privacy, transparency and unintended consequences. Current trends in learning analytics.

Readings *(before class)*

1. Handbook of Learning Analytics, Ch.4, Ethics and Learning Analytics: Charting the (Un)Charted
2. Watch keynote address at LAK 2018 conference by Neil Selwyn

The End

May 6, 2019 – No Class (CHI conference)