EDUC 623: Big Data, Education, and Society
Fall 2021
Professor Ryan S. Baker

SYLLABUS

Instructor Info
Email: rybaker@upenn.edu
Office: GSE 439
Office hours: Wednesdays, 4pm-5pm and by appointment
Course time: Fridays, 830am-1020am
Office hours and course location: https://bluejeans.com/rybaker/
Class discussion forum: http://piazza.com/upenn/fall2021/educ623

Required Texts:
• None

Information on how to obtain course readings will be provided in class.

Course Goals: The growth of learning analytics and educational data mining has been met with both optimism and concern. Excitement about the possibilities of individualized, personalized, adaptive learning have emerged. But concerns that student privacy will be jeopardized, and that student futures will be forever shaped by data from long ago – or warped by an errant prediction about the student years into the future – have emerged as well.

In this class, we will discuss what learning analytics can do, what it has the potential to do for good, and what the potential is for harm. We will discuss multiple uses and applications of analytics, where simple steps can mitigate risk, the relationship between validity and risk, and where risk mitigation will do more harm than good. We will do so in the context of real-world educational systems, challenges, problems, and with reference to original sources as much as possible.

Course Pre-requisites: None, but some prior experience with statistics or data mining recommended.

Assignments:

This class will have one primary assignment with multiple sub-assignments.

In this project, students will propose a learning analytics application in a group. In the first sub-project, due on October 4, you will propose the application and discuss past related work (in both research and practice). In the second sub-project, due on October 25, you will perform a needs assessment targeted towards articulating what societal or educational need the application addresses. In the third sub-project, due on November 15, you will discuss the risks and challenges inherent in their solution and how they can be mitigated. In the fourth sub-project, due on December 17, you will present your project as if they were submitting it to a potential funder. Part of your grade on each of these sub-projects will be commenting on other groups' submissions.

Extensions for the assignments will only be available in case of instructor error or extreme circumstances (assignments in other classes, research studies, and so on do not count as extreme circumstances; serious injury, illness, or death in the family do count as extreme circumstances). Outside of these circumstances, late hand-ins will not be accepted (e.g. zero credit will be given).

No examinations will be given in this class.

Class participation involves both attendance and active (and constructive) participation in classroom discussions and on the discussion forum (beyond participation as required for the assignments). While it is not expected that you will memorize every paper assigned for the class, it is expected that you will have studied the readings to the degree that you can participate actively in discussions.

Grading
- Project Proposal   20%
- Needs Assessment   20%
- Risks and Challenges   20%
- Final Project   20%
- Class Participation   20%

Course Schedule
Big Data, Education, and Society
Professor Ryan S. Baker

Fri, Sep. 3
The Emerging Era of Big Data in Education

Readings

- None

Fri, Sep. 10
What is Learning Analytics?

Readings

- Baker, R., Siemens, G. (2022) Educational data mining and learning analytics. To appear in Sawyer, K. (Ed.) *Cambridge Handbook of the Learning Sciences: 3rd Edition*.
- Wise, A. F. (2019). Learning Analytics: Using Data-Informed Decision-Making to Improve Teaching and Learning. In *Contemporary Technologies in Education* (pp. 119-143). Palgrave Macmillan, Cham.

Fri, Sep. 17
At-Risk Prediction

Readings

- Milliron, M. D., Malcolm, L., & Kil, D. (2014). Insight and action analytics: Three case studies to consider. *Research & Practice in Assessment*, *9*.
- Dawson, S., Jovanovic, J., Gašević, D., & Pardo, A. (2017). From prediction to impact: Evaluation of a learning analytics retention program. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 474-478). ACM.
- Coleman, C., Baker, R., Stephenson, S. (2019) A Better Cold-Start for Early Prediction of Student At-Risk Status in New School Districts. *Proceedings of the 12th International Conference on Educational Data Mining*, 732-737.
- Christie, S. T., Jarratt, D. C., Olson, L. A., & Taijala, T. T. (2019). Machine-Learned School Dropout Early Warning at Scale. *Proceedings of the 12th International Conference on Educational Data Mining*.

Fri, Sep. 24
Reports for School Personnel

Readings

- Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the assistment system. *Technology Instruction Cognition and Learning*, *3*(1/2), 63
- Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-Designing a Real-Time Classroom Orchestration Tool to Support Teacher-AI Complementarity. *Journal of Learning Analytics*, *6*(2), 27-52.
- Wise, A. F., & Jung, Y. (2019). Teaching with Analytics: Towards a Situated Model of Instructional Decision-Making. *Journal of Learning Analytics*, *6*(2), 53-69.
- Adair, A., & Dickler, R. (2020). Supporting Teachers Supporting Students: Iterative Development of TIPS in a Teacher Dashboard. In *International Conference of the Learning Sciences* (Vol. 3).

Fri, Oct. 1
Reports for Parents

Readings

- Broderick, Z., O'Connor, C., Mulcahy, Heffernan, N. & Heffernan, C. (2011). Increasing Parent Engagement in Student Learning Using an Intelligent Tutoring System. *Journal of Interactive Learning Research*, 22(4)*, 523-550.
- Sousa, D. A., Luze, G., & Hughes-Belding, K. (2014). Preferences and attitudes toward progress reporting methods of parents from diverse backgrounds. *Journal of Research in Childhood Education*, *28*(4), 499-512.
- Bergman, P. (2021) Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment Investment. *Journal of Political Economy, 129* (1), 286-322.

Mon, Oct. 4
Assignments Due: Project Proposal

Fri, Oct. 8
Automated Intervention

Readings

- Corbett, A. (2001) Cognitive computer tutors: Solving the two-sigma problem. *UM2001, User Modeling: Proceedings of the Eighth International Conference*, 137–147.

- Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education*, *24*(4), 427-469

Fri, Oct. 15
Validity

Readings

- Baker, R.S. (2020) *Big Data and Education*. 6th Edition. Philadelphia, PA: University of Pennsylvania. Chapter 2-6.
- Mislevy, R. J. (2016). How developments in psychology and technology challenge validity argumentation. *Journal of Educational Measurement*, *53*(3), 265-292.
- Hand, D.J. (1998) Data Mining: Statistics and More? *The American Statistician, 52* (2), 112-118.
- Hand, D.J., Blunt, G., Kelly, M.G., Adams, N.M. (2000) Data Mining for Fun and Profit. *Statistical Science, 15* (2), 111-126.

Fri, Oct. 22
Rational Modeling and Model Validity

Readings

- Muldner, K., Burleson, W., Van de Sande, B., & VanLehn, K. (2011). An analysis of students' gaming behaviors in an intelligent tutoring system: predictors and impacts. *User modeling and user-adapted interaction*, *21*(1), 99-135
- Paquette, L., de Carvalho, A.M.J.A., Baker, R.S. (2014) Towards Understanding Expert Coding of Student Disengagement in Online Learning. *Proceedings of the 36th Annual Cognitive Science Conference*, 1126-1131.

Mon, Oct. 25
Assignments Due: Needs Assessment

Fri, Oct. 29
Generalizability

Readings

- Hawkins, D. M. (2004). The problem of overfitting. *Journal of chemical information and computer sciences*, *44*(1), 1-12.
- Baker, R.S. (2020) *Big Data and Education*. 6th Edition. Philadelphia, PA: University of Pennsylvania. Chapter 2-5.
- Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68-84

Fri, Nov. 5
Discrimination and the Perpetuation of Bias

Readings

- Garcia, M. (2016). Racist in the Machine: The Disturbing Implications of Algorithmic Bias. *World Policy Journal*, *33*(4), 111-117
- Loukina, A., Madnani, N., & Zechner, K. (2019). The many dimensions of algorithmic fairness in educational applications. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 1-10).
- Kizilcec, R. F., & Lee, H. (2020). Algorithmic fairness in education. *arXiv preprint arXiv:2007.05443*.
- Baker, R. S., & Hawn, A. (2021). Algorithmic Bias in Education. edarXiv preprint pbmvz. https://edarxiv.org/pbmvz/

Fri, Nov. 12
Implementation Fidelity

Readings

- Feng, M., Roschelle, J., Heffernan, N., Fairman, J., & Murphy, R. (2014). Implementation of an intelligent tutoring system for online homework support in an efficacy trial. *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp. 561-566).
- Wise, A. F., & Vytasek, J. (2017). Learning analytics implementation design. *Handbook of learning analytics*, *1*, 151-160.
- Bingham, A. J., Pane, J. F., Steiner, E. D., & Hamilton, L. S. (2018). Ahead of the curve: Implementation challenges in personalized learning school models. *Educational Policy*, *32*(3), 454-489.
- Phillips, A., Pane, J. F., Reumann-Moore, R., & Shenbanjo, O. (2020). Implementing an adaptive intelligent tutoring system as an instructional supplement. *Educational Technology Research and Development*, *68*(3), 1409-1437.

Mon, Nov. 15
Assignments Due: Risks and Challenges

Fri, Nov. 19
Student Privacy

Readings

- Sabourin, J., Kosturko, L., FitzGerald, C., & McQuiggan, S. (2015). Student Privacy and Educational Data Mining: Perspectives from Industry. *Proceedings of the International Conference on Educational Data Mining*.
- Arnold, K.E., Sclater, N. (2017). Student Perceptions of their Privacy in Learning Analytics Applications. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference.*

- Klose, M., Desai, V., Song, Y., & Gehringer, E. (2020). EDM and Privacy: Ethics and Legalities of Data Collection, Usage, and Storage. *Proceedings of the International Conference on Educational Data Mining*.

Wed, Nov. 24 *DIFFERENT DAY OF THE WEEK BY GSE POLICY*
<u>Interpretability, Explainability, and Transparency</u>

Readings

- Hlosta, M., Papathoma, T., & Herodotou, C. (2020). Explaining Errors in Predictions of At-Risk Students in Distance Learning Education. In *International Conference on Artificial Intelligence in Education* (pp. 119-123).
- Zhou, T., Sheng, H., & Howley, I. (2020). Assessing Post-hoc Explainability of the BKT Algorithm. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 407-413).
- Liu, R., & Koedinger, K. R. (2017). Going beyond better data prediction to create explanatory models of educational data. *The Handbook of learning analytics*, 69-76.

Fri, Dec. 3
<u>Beneficence</u>

Readings

- Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room: the obligation to act. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference.*
- Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology*.

Fri, Dec. 10
<u>Big Data, Big Science, and Longitudinal Follow-up</u>

Readings

- Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, *24*(4), 470-497.
- Andres, J.M.L., Baker, R.S., Gasevic, D., Siemens, G., Crossley, S.A., Joksimovic, S.(2018) Studying MOOC Completion at Scale Using the MOOC Replication Framework. In *Proceedings of the International Conference on Learning Analytics and Knowledge*, 71-78.
- Fyfe, E., de Leeuw, J., Carvalho, P., Goldstone, R., et al. (2019). ManyClasses 1: Assessing the generalizable effect of immediate versus delayed feedback across many college classes. psyarXiv paper 4mvyh .
- Patikorn, T., Baker, R.S., Heffernan, N.T. (2020) ASSISTments Longitudinal Data Mining Competition Special Issue: A Preface. *Journal of Educational Data Mining, 12* (2), i-xi.

Fri, Dec. 17
Class Presentations of Assignment 4: Final Project

Class will run an extra hour long this session