MLND Capstone Project Description - Education

Education Capstone Project

The following are a set of available materials, datasets, and questions to get you started thinking about how to apply machine learning to better understanding education.

Available Courses

- Big Data in Education (edX)
- Big Data and Education (Columbia)
- Educational Technology (Georgia Tech and Udacity)

Available Datasets

The following datasets are available in a number of different formats, including CSV, and can be processed by the tools and techniques of your choice that you have learned thus far in this Nanodegree program.

- Education Data Analysis Tool, from NCES
- Elementary/Secondary Information System, from NCES
- International Data Explorer, from NCES
- Integrated Postsecondary Education Data System Delta Cost Project Database, from NCES
- National Assessment of Educational Progress Data Explorer, from NCES
- College Scorecard, from Data.gov
- Consolidated State Performance Report, from Data.gov
- PSLC DataShop, from Carnegie-Mellon University
- Educational Data Mining Challenge, from Carnegie-Mellon University

You are also welcome to find additional educational datasets and pursue your own research.

Sample Questions

The following are some questions you may answer with the datasets above. You are also free and encouraged to develop your own questions based on the datasets in which you are interested; these are just intended to get you started.

- What university-level factors predict the presence of a strong retention and graduation rate? (Answerable with the College Scorecard)
- What types of financial aid are most effective in maximizing completion rates, minimizing student loan burdens, and maintaining well-funded schools? (Answerable with the Integrated Postsecondary Education Data System Delta Cost Project Database)
- What determines states' degree of compliance with NCLB standards, and to what extent does that compliance improve educational outcomes? (Answerable with the Consolidated State Performance Report)
- What types of usage patterns exist during students' interaction with intelligent tutoring systems, and what usage patterns correlate with ultimate successful learning? (Answerable with the Assistments dataset from PSLC DataShop)
- How does past success predict future success? Can we predict the need for educational interventions early enough to be successful? (Answerable with the dataset from the 2010 Educational Data Mining Challenge)

Relevant Sources

- Realizing the Opportunity for Big Data in Education, by Aubrey Francisco
- Big Data and Education: What's the Big Idea?, by Oliver Marsh, Lajos Maurovich-Horvat, and Olivia Stevenson
- The Coming Big Data Education Revolution, by Doug Guthrie
- Big Data Enters the Classroom, by Lisa Fleisher