Just-in-Time Intervention

Hosuk Choi, Jianan Jin, Anusha Prasad, Manley Roberts, Ania Thomas, Matt Yang

Meet the Team

Manley Roberts
CS

Hosuk Choi
CS

Jianan Jin
Analytics

Ania Thomas
CM

Anusha Prasad
CS

Matt Yang
CM

High-Level Motivation

- A key component to our projects success is PREVENTION
  - We want to intervene prior to a students predicted drop out date/ period
- Potential ways to communicate our predictions in the most efficient and effective way
  - Without discouraging the student
  - Without providing the professor with too much of an extra workload
- Two options we're currently considering
  - Pop-up alerts
    - Inspired by Coursera
  - Email correspondence on a specific day

Subteam History

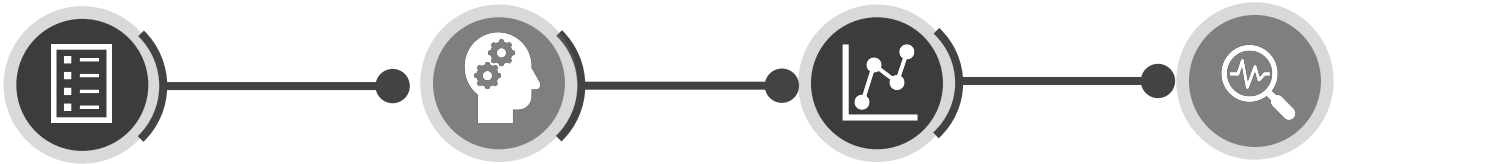
- **Spring 2019 - ‘Research Subteam’ Research Goal**
  - To use various subsets of course data related to performance in a course (ISyE 6501, Introduction to Analytics Modeling), combined with various machine learning/statistical models, to predict final success in the course.
  - Once these models are trained, we will attempt to make conclusions about **particular predictors** of **success** or **failure** in the course.
- **Research Scope**
  - 1 semester of a single class, only for verified students (n ~= 300)
  - Predictions made based on clickstream and assignment data
  - Not too much focus on selecting **only the most important features**

Data Sources

- Clickstream data from edX:
  - Past data utilized:
    - Starts and stops on course videos
    - Clicks on next sections for the course
  - Still determining other possibilities for clickstream data within edX (will be elaborated through data exploration)

- Assignment data:
  - Examining assignment grade data for every week of the course
  - Utilizing lag data (i.e. grade data from 3 weeks ago, 2 weeks ago, and 1 week ago)

Project Flow



Data Cleaning and Exploring

Feature Engineering

Model Training

Parameter Tuning and Model Evaluation

Modeling

Different Aspects of the problem that we can attempt to predict with a model

- **How long until drop out?**
- **Will the student drop out next week?**
  - What is the probability of the student dropping out the next week?
- **At a given moment in time, predicting a students' final grade/success**
  - What is the probability the student will finish the course?

Modeling cont.

- **Machine Learning**
  - Machine learning uses computers to predict unknown object attributes through the recognition of patterns in data.
- **Supervised Learning**
  - We have a lot of various types of data (measured attribute values)
    i. Specifically, each student's data would come in a tuple
  - Using SOME of the data, we can create a function that maps inputs onto outputs
    i. E.g. Quiz performance onto final grade
  - Then using a DIFFERENT subset of the data, we can test our function's accuracy
  - And repeat...

Modeling cont.

Survival Analysis

- Cox Proportional-hazard Model
- Outcome variable of interest is time until a dropout occurs
- To examine how specific factors influence the rate of a dropout happening at a particular week

Data Arrangements

- Use all the past data (excluding the most recent semester) to train our model and test that model using the most recent semester (hence, why it was excluded)
- Use a mix of all data and use an amount of it to create and train the model and then test it using the remainder of the data
- We'll do both data arrangements and see how each one goes and then use that for advancement in the next semester

Next Steps

- Divide team into two sub-subteams
  - MongoDB team - focuses on acquiring and aggregating data from MongoDB clickstream data
  - SQL team - focuses on acquiring and aggregating data from SQL assignment server
- Explore data, engineer potentially helpful attributes from existing attributes
- Develop pipelines to acquire data, transform data, train models, and evaluate success
- Examine whether it is reasonable to predict student success using cross-semester models
- **Long term** - Enable predictions for current semesters of the class.