



Predicting Dropout

Predicting Student Dropout Risk in MOOC courses
Capstone project by Dmytro Kovalchuk
(DRAFT)

LearnMoodle



learnmoodle.com, August 2016

Total Enrollment	6119	100%
Completed “May we use your data for research?” survey (required to access remainder of course)	2769	45%
Granted permission to use data in research	2167	35%
Earned “Learn Moodle participant Aug 2016” badge (requires a post in “Forum – Introduce yourself!”)	1566	26%
Earned “Learn Moodle completer Aug 2016” (requires completion of all activities in course)	735	12%

Objectives

- Build predictive model to help teachers identify users that are at early risk of dropout
- Design data-pipeline and model framework that is suitable for product integration

Data

Dataset

- Users
- Standard Logs
- Course Modules
- Completions
- Grades
- Badges

Features

Did the user complete this week LOs by the deadline?

- wk1_page
- wk1_book
- wk1_forum
- wk1_quiz

Dependent Variable

Did the user complete all required LOs at the end of the course?

Model

Logistic Regression:

```
class_weight='balanced'  
C=0.1  
penalty='l1'
```

```
205  
206 # -----  
207 # ModelFactory: Week 1 - all L0s  
208 # -----  
209  
210 # Scores:  
211           precision    recall  f1-score   support  
212  
213      False      0.88      0.77      0.82      405  
214       True      0.50      0.68      0.58      137  
215  
216  avg / total      0.78      0.75      0.76      542  
217  
218 # Coefficients:  
219 1_forum : 0.994836986974  
220 1_book_1 : 0.683228744567  
221 1_quiz : 0.673524874656  
222 1_feedback : 0.617447471855  
223 3_data : -0.452692961904  
224 0_page_3 : 0.285716297419  
225 1_book_2 : -0.24369266246  
226 1_choice : 0.103738686854  
227 1_page_1 : -0.0687224891645  
228 0_forum : 0.0661934796414  
229
```

Data Pipeline

DataExplorer

Loads Raw Data

Cleaning & Munging

EDA functions

Saves Interim Data

Saves Processed Data

FeatureFactory

Load Processed Data

Create Features

Save Features

Dependent Variable

ModelFactory

Trains, Stores and Loads Models

- WK1_model
- WK2_model
- WK3_model
- WK4_model
- others

- **Questions?**
- **Thank you**