



# Sign up, Tune in, don't Drop out

---

Predicting student dropout risk in MOOCs  
by Dmytro Kovalchuk

[github.com/KDmytro/moodle](https://github.com/KDmytro/moodle)  
[linkedin.com/in/KDmytro](https://linkedin.com/in/KDmytro)

# Business Problem

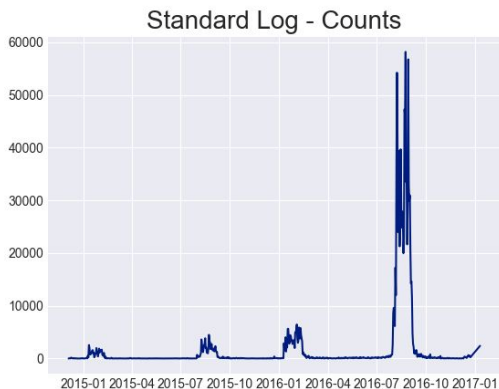
- Who is at risk of dropout?
- What are the predictive factors?
- Early intervention is effective

Whitehill J, Williams J, Lopez G, Coleman C, Reich J. Beyond Prediction: First Steps Toward Automatic Intervention in MOOC Student Stopout, in Educational Data Mining. ; 2015.

- Limited tools available for instructors to automate intervention

# LearnMoodle.com

- August 2016 cohort, 4 weeks duration
- 6119 enrolled students
- Anonymized dataset extract directly from moodle SQL DB
- Standard log file: 1.3 mil records



ABOUT · LEARN MOODLE MOBILE APPYou are not logged in. (Log in)

## Inspiring better teaching everywhere

Welcome to the official Moodle MOOC site. Our next course will be starting on Monday 19 June 2017.

[SIGN UP](#)[LOG IN](#)

### Teaching with Moodle

This free course is designed for anybody who wants to use the Moodle learning platform for teaching, whether it be in a school, a university, a company or just personal interest!

The four-week course is also a great opportunity to connect with the vibrant Moodle community dedicated to sharing resources, ideas and anything that could help inspire better teaching practices everywhere.





#### Overview

An activity-based learning course for teachers new to Moodle. Four weeks, 3 to 4 hours a week. No fees.



#### Earn badges

As a Mozilla Open Badges issuer, badges will be awarded for active participation and upon course completion!



#### Connect

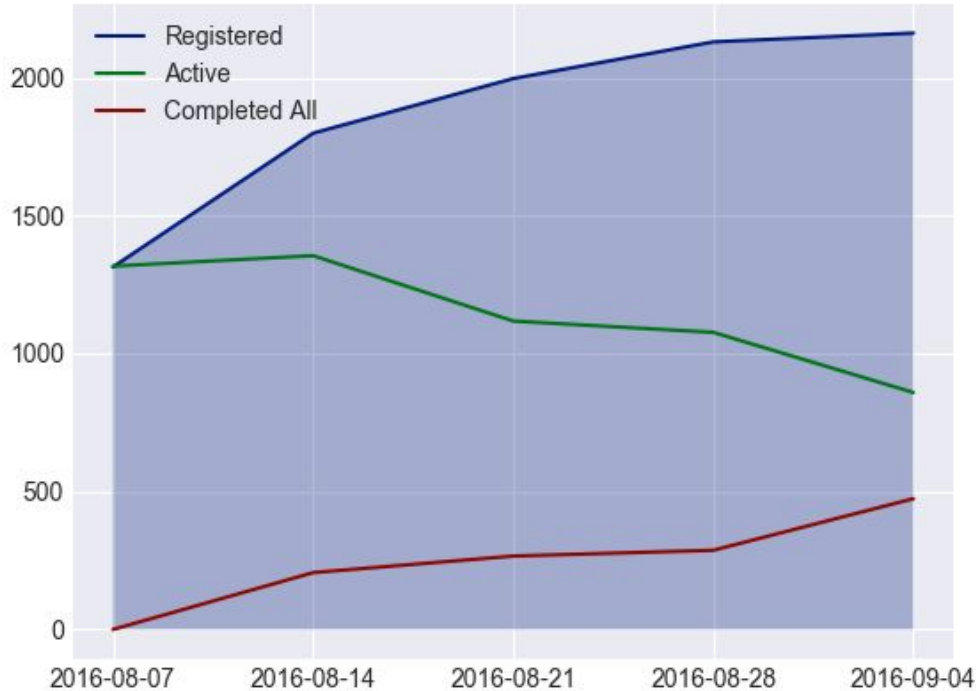
Join the conversation on Twitter using #learnmoodle, and engage with other teachers in the course forum.



#### Share this page

We love teachers and want to inspire better teaching, so help us spread the word to educators everywhere!

# Student Activity Pattern



- Continuous inflow of new students
- High completion activity before the end of the course

# Feature Selection

## Engagement

Total count of activities within a given timeframe

- Number of logins
- Days since last login
- Page views
- Forum posts
- Messages sent
- Prior registrations

## Completions

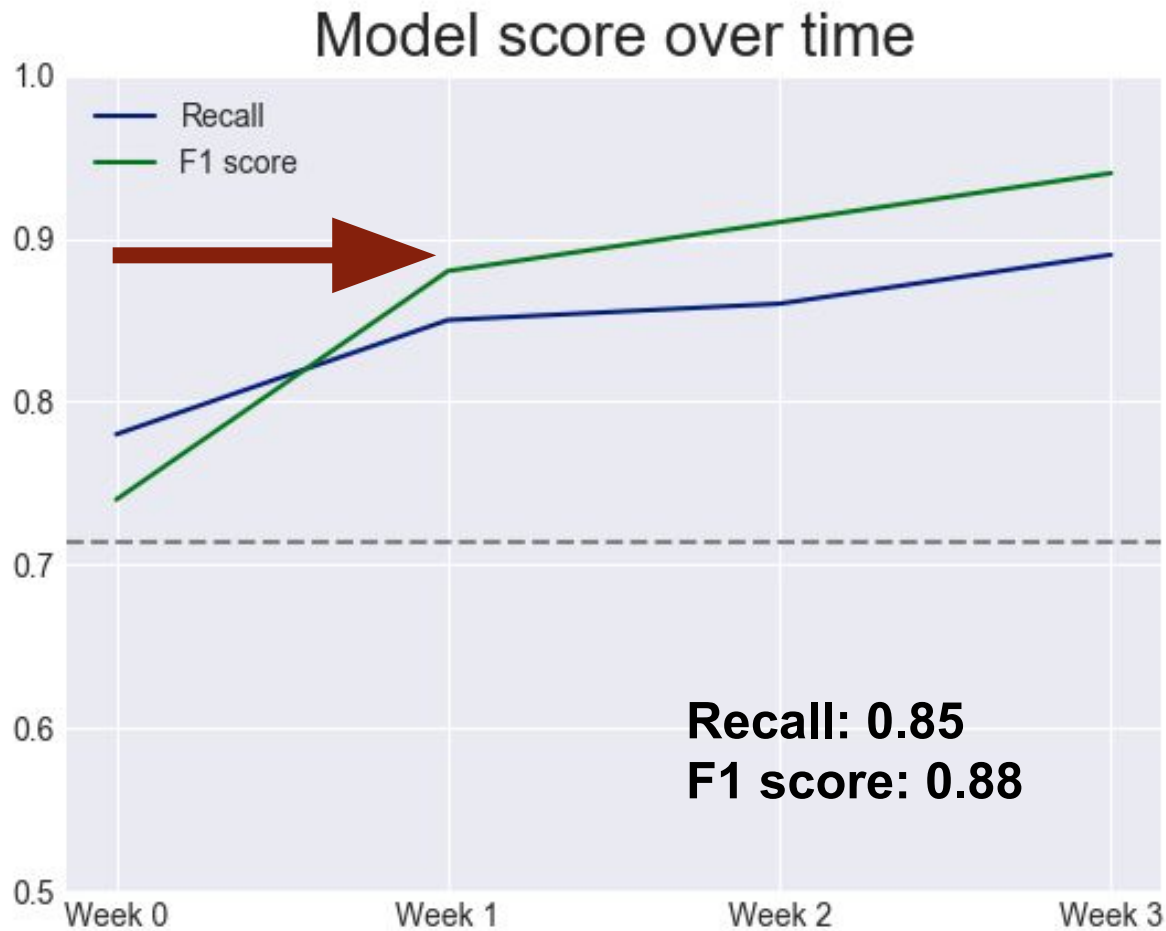
Did the user complete this week learning objects by the deadline?

- WK1 - Page
- WK1 - Forum
- WK1 - Quiz

## Dependent Variable

Did the user complete all required learning objects at the end of the course?

# Model



# Model

- Page views
- Days since last login
- Feedback
- Wk1 - Forum

Model score over time



# Model

- Page views
- Number of logins
- Posts created
- Wk2 - Quiz
- Wk2 - Glossary
- Wk2 - Wiki
- Wk3 - Quiz
- Wk3 - Forum





# Data Pipeline

DataExplorer

**SQL Interface**

**Cleaning & Preprocessing**

**EDA functions**

FeatureFactory

**Generates Features**

**Dependent Variables**

ModelFactory

**Trains, Stores and Loads  
Models**

- WK1\_model
- WK2\_model
- WK3\_model
- WK4\_model

# Next steps

- Collect more data from moodle:
  - To test model accuracy within other cohorts
  - To generalize the model across different courses
- Develop and implement PHP plug-in for Moodle

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

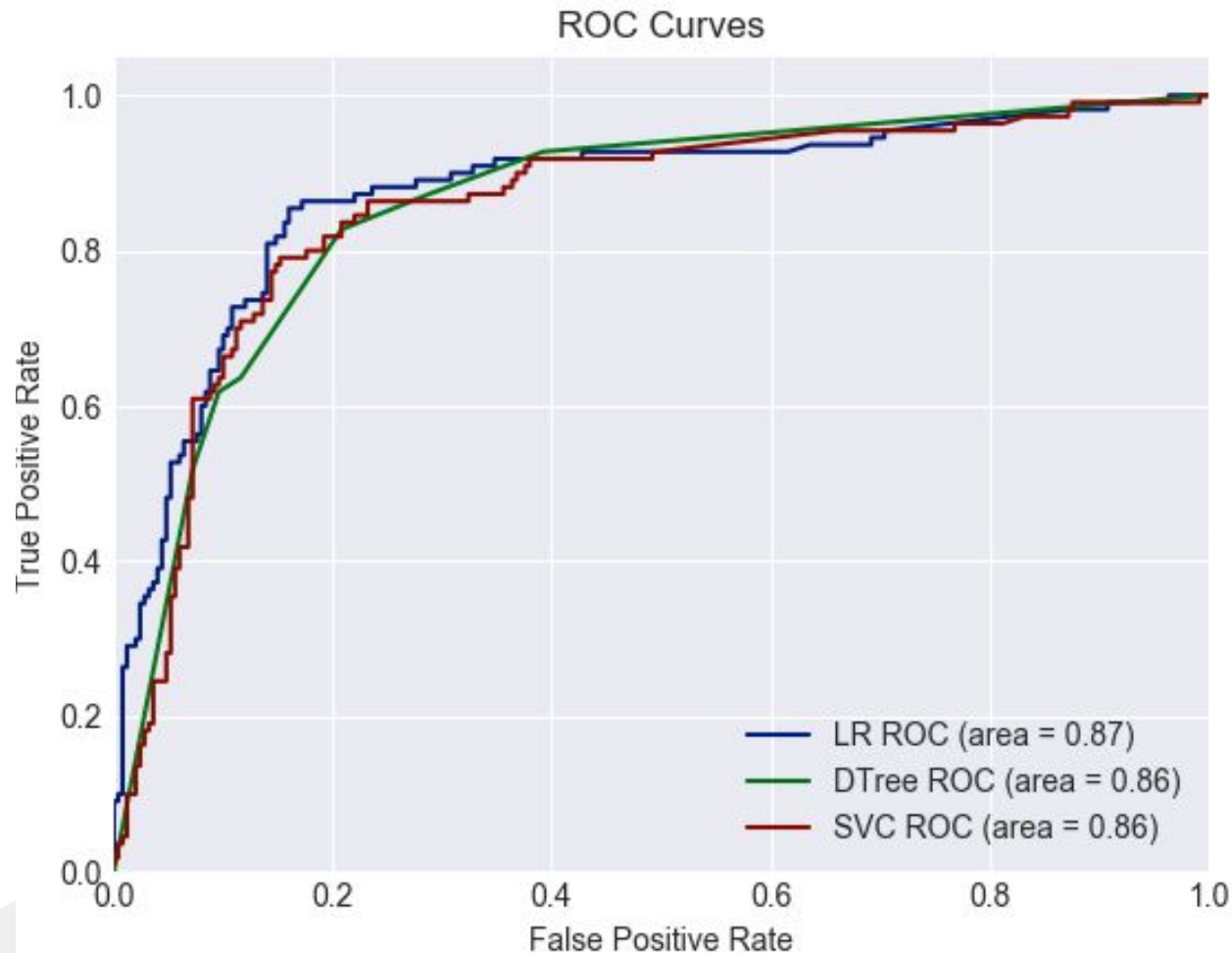
Dmytro Kovalchuk  
[github.com/KDmytro/moodle](https://github.com/KDmytro/moodle)  
[linkedin.com/in/KDmytro](https://linkedin.com/in/KDmytro)

# Appendix



# Model Selection

- Logistic Regression
- Decision Tree
- SVC





# Week 3 Model

## Logit Regression Results

Dep. Variable:	y	No. Observations:	2130			
Model:	Logit	Df Residuals:	2100			
Method:	MLE	Df Model:	29			
Date:	Tue, 23 May 2017	Pseudo R-squ.:	0.5303			
Time:	08:23:12	Log-Likelihood:	-602.56			
converged:	True	LL-Null:	-1282.9			
		LLR p-value:	2.761e-268			
	coef	std err	z	P> z	[0.025	0.975]
n_logins	0.6791	0.194	3.498	0.000	0.299	1.060
n_viewed	0.4724	0.206	2.291	0.022	0.068	0.877
n_posts_created	2.3156	0.721	3.210	0.001	0.902	3.729
last_login	-0.1240	0.114	-1.085	0.278	-0.348	0.100
n_msg_sent	-0.7332	0.510	-1.439	0.150	-1.732	0.265
n_prior_enrollment	0.0527	0.125	0.421	0.673	-0.192	0.298
0_forum	-0.0244	0.116	-0.210	0.834	-0.253	0.204
0_page_1	-0.0611	0.161	-0.379	0.705	-0.367	0.255
0_page_2	-0.0298	0.155	-0.193	0.847	-0.333	0.273
0_page_3	0.1626	0.118	1.381	0.167	-0.068	0.393
1_forum	-0.2203	0.115	-1.916	0.055	-0.446	0.005
1_book_1	-0.0895	0.152	-0.588	0.556	-0.388	0.209
1_book_2	-0.0495	0.174	-0.285	0.776	-0.390	0.291
1_quiz	0.0197	0.163	0.121	0.904	-0.300	0.340
1_feedback	0.2555	0.154	1.659	0.097	-0.046	0.557
1_choice	-0.0980	0.124	-0.792	0.428	-0.341	0.145
1_page_2	-0.0785	0.163	-0.483	0.629	-0.397	0.240
1_page_1	0.0022	0.101	0.022	0.983	-0.197	0.201
2_quiz	0.3487	0.146	2.388	0.017	0.063	0.635
2_book	-0.1665	0.159	-1.050	0.294	-0.477	0.144
2_glossary	0.5034	0.152	3.310	0.001	0.205	0.802
2_survey	0.0275	0.207	0.133	0.894	-0.378	0.433
2_wiki	0.4222	0.189	2.235	0.025	0.052	0.792
2_page_1	-0.1180	0.148	-0.800	0.424	-0.407	0.171
3_quiz	0.6583	0.156	4.221	0.000	0.353	0.964
3_book	-0.1175	0.160	-0.734	0.463	-0.431	0.196
3_forum	1.2086	0.227	5.313	0.000	0.763	1.654
3_data	-0.0188	0.126	-0.149	0.882	-0.266	0.228
3_page_1	0.1976	0.111	1.782	0.075	-0.020	0.415
3_page_2	0.2748	0.199	1.378	0.168	-0.116	0.666