Applying Predictive Models to Course Curricula for Early Prediction of Struggling Students

UNIVERSITY OF WISCONSIN PLATTEVILLE

Nathanael Braukhoff, Austin FitzGerald, Evan Majerus, Zhiwei (Henry) Yang (University of Wisconsin - Platteville)

Faculty Mentors: Douglas Selent (University of Wisconsin - Platteville) & Seth Adjei (Northern Kentucky University)

Abstract

The purpose of this research is to provide data and tools necessary for students and faculty advisors to predict and prevent academic struggle.

Research Questions

1. What is the percentage of students who have struggled while attending UW-Platteville?
2. How well can we predict a student's next term GPA using their previous term GPA?
3. Given varying amounts of prior course performance data, what is the probability that a student will graduate?
4. Given student performance data on prerequisite courses, how accurate and how far into the future can we predict post-requisite course performance?

Data

Our dataset consists of the historical grade data of 689 unique students from University of Wisconsin - Platteville graduates and withdrawals between the years 2013 - 2018.

Data Sample

Anonymized student ID	Year	Course Name	Grade	Academic Standing	Course Credits
9535	2007	World Population, Food and Resources	B	Good	3
2197	2007	Global Business	D	Probation	3
5447	2007	Leadership and Management	C	Good	3

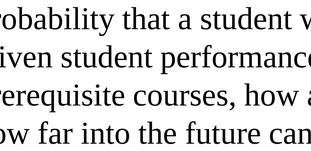
Methodology

Models

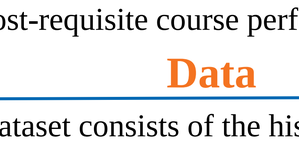
- Gradient Boosted Trees
- Linear/Logistic Regression
- Neural Network (For next term GPA prediction)
- ZeroR (Baseline)
- Bayesian Network (Future work)

Experiments

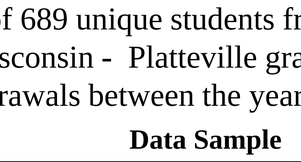
- Predicting next term GPA with data from the previous term
- Predicting if a student will graduate or not with varying amounts of term GPAs
- Predicting future course grades
  - All prerequisite courses



  - Immediate prerequisite courses



  - Root prerequisite courses

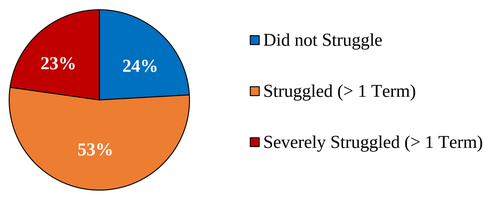


Evaluation

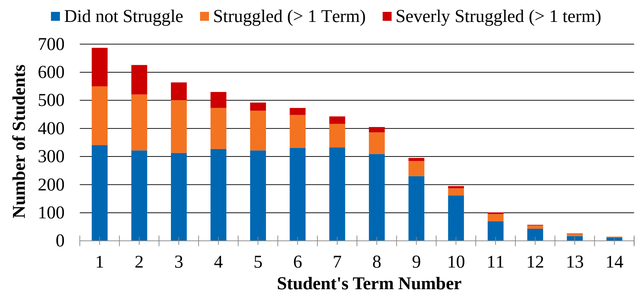
- Next Term GPA: We used five-fold stratified (on term number) student-level cross validation
- Graduation: Five-fold stratified (on binary graduation status) student-level cross validation.
- Future Course Grade: We used five-fold stratified (on grade) student-level cross validation
- We evaluated our models using Normalized Root Mean Square Error (NRMSE)

Analysis & Results

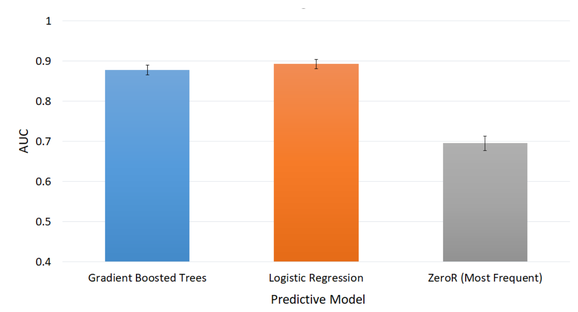
Students Who Struggled



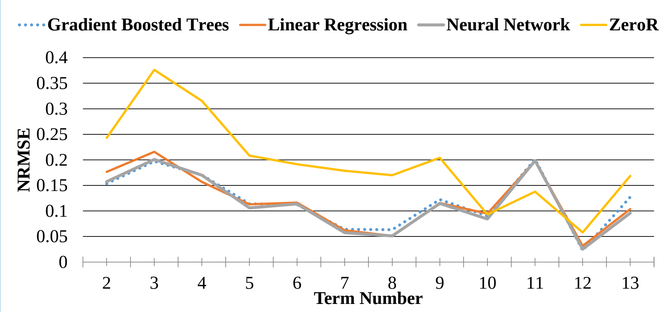
Students Who Struggled Per Term



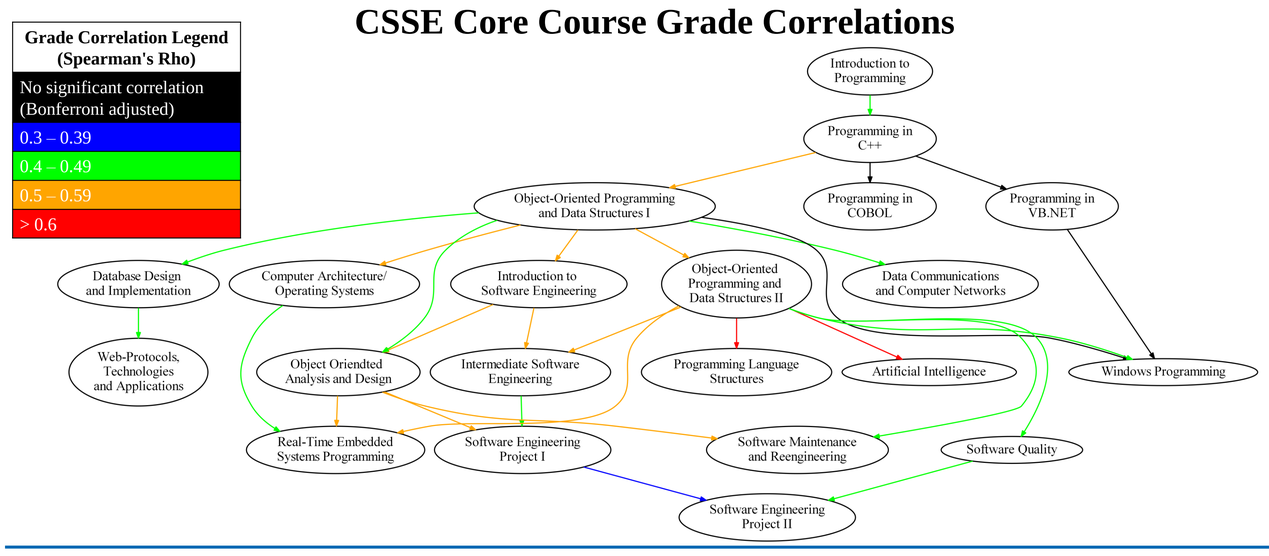
Graduation Status Prediction - Model Performance



Next Term GPA Prediction Model Performance



CSSE Core Course Grade Correlations



Post-requisite Grade Prediction - Model Performance

Accuracy / Within Half-Grade Accuracy / Within Full Grade Accuracy

	All Prerequisite Courses	Immediate Prerequisite Courses	Root Prerequisite Courses
Gradient Boosted Trees	86.2% / 92.8% / 99.1% NRMSE = 0.09	33.0% / 51.3% / 84.7% NRMSE = 0.28	82.7% / 91.1% / 98.5% NRMSE = 0.10
Logistic Regression	51.5% / 76.6% / 98.0% NRMSE = 0.14	54.6% / 79.9% / 98.8% NRMSE = 0.13	52.6% / 78.5% / 98.6% NRMSE = 0.15
ZeroR - Mean	13.5% / 40.9% / 78.9% NRMSE = 0.28	13.4% / 40.9% / 78.9% NRMSE = 0.27	15.4% / 44.0% / 80.5% NRMSE = 0.27

Post-requisite Grade Prediction Confusion Matrices

All Prerequisite Courses

Gradient Boosted Trees

Actual Grade \ Predicted Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	F
A	943	65	7	21	0	0	1	0	0	0	0
A-	67	435	34	21	1	0	0	0	0	0	0
B+	5	13	487	63	6	1	2	0	0	0	0
B	7	1	6	1468	5	2	2	1	0	0	0
B-	5	2	2	23	545	15	12	0	0	1	0
C+	1	0	3	17	24	395	28	1	0	2	1
C	1	0	1	8	3	9	927	14	1	3	9
C-	0	0	0	2	2	1	65	229	3	17	36
D+	0	0	0	0	0	1	26	18	7	8	12
D	0	0	0	1	2	2	28	40	2	177	64
F	0	0	0	3	0	0	30	43	4	35	407
Percent Correct	91.6	84.3	90.2	90.2	92.7	92.7	82.7	66.2	41.2	72.8	76.9
Percent Within Whole Grade	99.3	100.0	99.8	99.6	99.7	99.5	97.1	99.7	100.0	98.8	89.0

Immediate Prerequisite Courses

Gradient Boosted Trees

Actual Grade \ Predicted Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	F
A	647	24	10	278	5	2	51	5	0	1	7
A-	198	22	10	246	16	1	47	4	0	3	9
B+	141	15	16	304	18	3	54	3	0	5	14
B	193	17	27	922	29	14	214	16	1	4	46
B-	41	2	3	354	38	7	107	14	0	4	30
C+	27	7	2	231	15	12	125	13	1	8	25
C	41	7	8	385	22	13	357	16	0	21	97
C-	5	1	3	107	10	11	123	29	0	9	53
D+	2	0	1	26	2	2	27	4	0	1	6
D	6	0	3	67	3	5	100	17	0	34	77
F	10	0	1	99	8	4	135	14	1	29	209
Percent Correct	49.4	23.2	19.0	30.5	22.9	16.2	26.6	21.5	0	28.6	36.5
Percent Within Whole Grade	89.9	84.2	81.0	90.1	89.2	83.8	78.6	79.3	66.7	79.0	49.9

Root Prerequisite Courses

Gradient Boosted Trees

Actual Grade \ Predicted Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	F
A	296	20	10	17	1	0	0	0	0	0	0
A-	24	164	16	15	0	0	0	0	0	0	0
B+	5	12	158	28	4	0	0	0	0	0	0
B	0	0	3	520	4	0	2	0	0	0	1
B-	1	0	0	11	177	9	5	1	0	2	0
C+	0	0	0	3	10	9	116	12	0	0	1
C	0	0	0	0	4	2	7	306	3	0	3	7
C-	0	0	0	0	1	0	34	49	1	2	2
D+	0	0	0	0	0	0	9	1	0	2	6
D	0	0	0	0	1	2	21	8	0	32	19
F	0	0	0	2	0	1	15	6	3	14	103
Percent Correct	90.8	83.7	83.2	85.7	88.9	85.9	75.7	72.1	0.0	58.2	74.1
Percent Within Whole Grade	99.7	100.0	100.0	99.7	99.0	97.8	96.3	100.0	100.0	96.4	87.8

All Prerequisite Courses

Logistic Regression

Actual Grade \ Predicted Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	F
A	984	29	10	12	0	1	1	0	0	0	0
A-	49	144	67	278	12	6	0	2	0	0	0
B+	7	58	77	362	40	10	18	3	1	1	0
B	6	52	98	1053	129	47	85	15	5	1	1
B-	1	7	27	348	112	32	64	8	3	3	0
C+	0	1	10	175	66	43	147	22	3	4	1
C	1	1	11	191	73	61	540	68	11	19	0
C-	0	0	1	29	18	21	189	64	10	22	1
D+	0	0	0	2	0	4	38	15	4	8	1
D	0	0	0	1	4	6	115	33	19	115	23
F	0	0	1	1	1	1	10	7	10	34	457
Percent Correct	93.9	49.3	25.5	42.9	24.6	18.5	44.7	27.0	6.1	55.6	94.4
Percent Within Whole Grade	99.8	99.3	95.7	98.7	98.9	94.0	97.6	91.6	86.4	95.7	99.2

Immediate Prerequisite Courses

Logistic Regression

Actual Grade \ Predicted Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	F
A	1009	12	4	3	0	0	1	0	1	0	0
A-	33	168	70	279	2	2	2	0	0	0	0
B+	6	49	85	396	22	9	5	0	0	1	0
B	1	34	76	1153	107	29	73	7	3	0	0
B-	0	5	16	380	98	19	73	6	3	0	0
C+	0	0	2	194	59	21	177	9	2	2	0
C	0	0	4	183	55	66	590	51	10	8	0
C-	0	0	1	35	14	13	216	51	5	16	0
D+	0	0	0	2	0	2	44	9	2	12	0
D	0	0	0	2	3	6	122	30	10	131	8
F	0	0	0	1	1	0	4	7	3	22	472
Percent Correct	96.2	62.7	32.9	43.9	27.1	12.6	45.1	30.0	5.1	68.2	98.3
Percent Within Whole Grade	100.0	100.0	98.1	98.5	98.9	95.2	99.1	95.9	82.1	98.4	100.0

Root Prerequisite Courses

Logistic Regression

Actual Grade \ Predicted Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	F
A	317	16	6	5	0	0	0	0	0	0	0
A-	11	87	37	83	0	1	0	0	0	0	0
B+	3	30	36	126	11	1	0	0	0	0	0
B	1	20	43	393	36	9	23	3	1	1	0
B-	0	3	8	112	45	10	25	2	1	0	0
C+	0	2	4	53	28	8	54	2	0	0	0
C	0	0	2	77	27	22	181	17	3	3	0
C-	0	0	1	11	4	2	53	10	1	6	1
D+	0	0	0	2	1	1	10	2	0	1	1
D	0	0	0	1	1	1	32	8	4	26	10
F	0	0	0	0	2	0	1	5	5	13	118
Percent Correct	95.5	55.1	26.3	45.5	29.0	14.5	47.8	20.4	0.0	52.0	90.8
Percent Within Whole Grade	100.0	98.7	97.8	98.4	97.4	96.4	99.7	93.9	86.7	98.0	98.5

All Prerequisite Courses

ZeroR – Mean

Actual Grade \ Predicted Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	F
A	4	88	102	292	353	150	38	9	1	0	0
A-	1	27	57	180	185	87	15	6	0	0	0
B+	1	26	49	188	207	78	15	13	0	0	0
B	1	42	49	355	621	290	87	46	1	0	0
B-	0	3	22	181	240	115	23	21	0	0	0
C+	0	2	12	100	195	111	35	16	1	0	0
C	0	3	17	143	366	255	122	70	0	0	0
C-	0	1	2	42	145	101	31	33	0	0	0
D+	0	0	1	15	33	14	7	2	0	0	0
D	0	0	2	23	89	76	59	65	2	0	0
F	0	2	4	33	140	138	114	90	1	0	0
Percent Correct	57.1	13.9	15.5	22.9	9.3	7.8	22.3	8.9	0		
Percent Within Whole Grade	100.0	95.9	91.8	92.7	76.1	68.1	66.7	80.1	66.7		

Immediate Prerequisite Courses

ZeroR – Mean

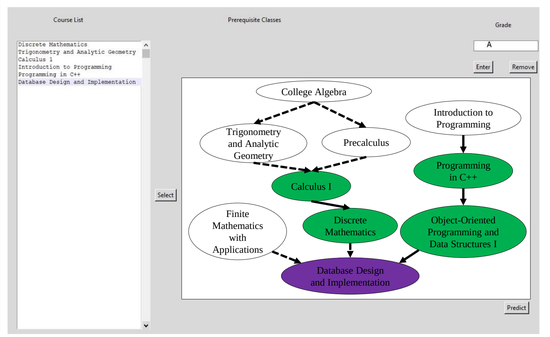
Actual Grade \ Predicted Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	F
A	4	88	102	285	353	152	36	9	1	0	0
A-	1	27	57	178	185	88	14	6	0	0	0
B+	1	26	49	185	206	79	14	13	0	0	0
B	1	42	49	349	618	292	85	46	1	0	0
B-	0	3	22	180	238	114	22	21	0	0	0
C+	0	2	12	97	192	113	33	16	1	0	0
C	0	3	17	139	364	256	118	70	0	0	0
C-	0	1	2	41	144	102	29	32	0	0	0
D+	0	0	1	15	32	15	6	2	0	0	0
D	0	0	2	22	89	77	55	65	2	0	0
F	0	2	4	31	135	140	109	88	1	0	0
Percent Correct	57.1	13.9	15.5	22.9	9.3	7.9	22.6	8.7	0.0		
Percent Within Whole Grade	100.0	95.9	91.8	92.8	76.2	68.0	66.8	79.9	66.7		

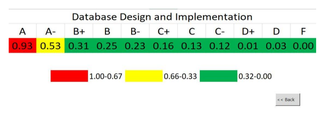
Root Prerequisite Courses

ZeroR – Mean

Actual Grade \ Predicted Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	F
A	4	37	56	86	126	19	16	0	0	0	0
A-	1	14	26	60	86	23	9	0	0	0	0
B+	1	12	27	51	83	19	14	0	0	0	0
B	1	6	17	90	270	67	79	0	0	0	0
B-	0	2	8	38	107	33	18	0	0	0	0
C+	0	3	4	28	61	36	19	0	0	0	0
C	0	3	4	38	120	87	80	0	0	0	0
C-	0	0	1	9	25	42	12	0	0	0	0
D+	0	0	0	4	10	2	2	0	0	0	0
D	0	0	2	5	33	27	16	0	0	0	0
F	0	2	1	12	52	23	54	0	0	0	0
Percent Correct	57.1	17.7	18.5	21.4	11.0	9.5	25.1				
Percent Within Whole Grade	100.0	89.9	94.5	92.9	77.3	75.7	70.8				

User Interface





Significance

- About 43% of the students have delayed their graduation to longer than 4 years.
- According to our data struggling is prevalent: 76% of the students have struggled at least once while attending UW-Platteville.
- Only 18.6% of students who fail a prerequisite course will continue to take the post-requisite.

Conclusions

- For predicting next term GPA all models performed significantly better than the baseline, but not significantly better than each other
- We can predict a student's graduation status approximately 86% accurately (within a 95% confidence interval).
- Predicting a future course grade can be done accurately with the right data. A GBT Classifier, with data from all prerequisites, can predict within a half-letter grade 92.8% of the time.

Acknowledgments

We would like to acknowledge funding from the following sources:

- The UW-Platteville Summer Undergraduate Scholars Program
- The UW-Platteville college of EMS
- UW-Platteville Scholarly Activity Improvement Fund
- The UW-Platteville Computer Science and Software Engineering department

Authors

Evan Majerus (majeruse@uwplatt.edu) is a sophomore at the University of Wisconsin - Platteville studying Software Engineering.

Nate Braukhoff (braukhoffna@uwplatt.edu) is a senior at the University of Wisconsin - Platteville studying Software Engineering.

Zhiwei Yang (yangz@uwplatt.edu) is a senior at the University of Wisconsin - Platteville studying Software Engineering.

Austin D. FitzGerald (fitzgeralaus@uwplatt.edu) is a sophomore at the University of Wisconsin - Platteville studying Software Engineering.

Dr. Douglas Selent (selentd@uwplatt.edu) is an Assistant Professor in the Computer Science and Software Engineering department at the University of Wisconsin - Platteville. https://sites.google.com/site/dougselent/

Dr. Seth Adjei (sadjeis1@nku.edu) is a Visiting Assistant Professor in the Computer Science department at Northern Kentucky University.