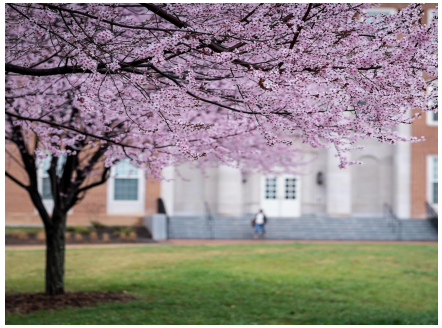
University Enrollment Case Study



Eagle Xuhui Ying

10/05/2022

Agenda

- Problem Statement
- Data Overview
- Exploratory Analysis
- Final Model
- Recommendations
- Appendix

Problem Statement

- **Our client**: University Administration
- **Situation**: We want to figure out factors that seem to be influencing a student's decision to enroll in the university and we need to give a description of the student who is most likely to attend
- **Goal**: Use classification modeling methods to predict whether or not a student will enroll in the university

Data

- 56,237 records, 29 columns
- Variables: campus visit code, in-state or not, referral contact count, self-initiated contact count, recruitment area, distance from university …

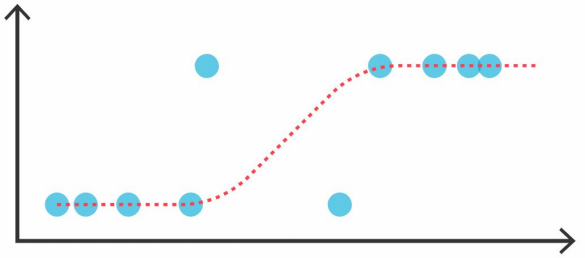
Logistic Regression Analysis

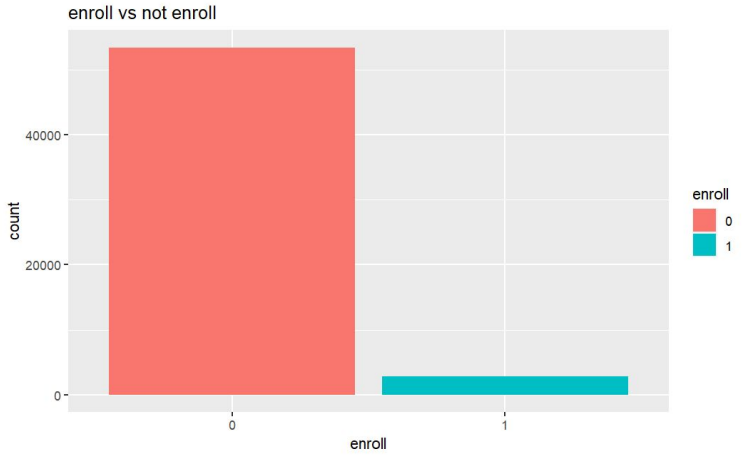
Explanatory Variables :

- Self-initiated contact count
- Time from first contact to enrollment date
- Campus visit code
- Number of indicated extracurricular interests
- ...

Response Variable:

- Enroll or not (0 or 1)





Data Cleansing

1. Reject 12 columns unsuitable to use admission decisions
2. Drop 2 columns contain > 20% missing values

Final Model

When 5-year enrollment rate from student's high school↑

5-year secondary interest code rate↑

5-year primary interest code rate↑ .....

The university enrollment rate will ↑

When solicited contact count↓

Time from first contact to enrollment date↓

The university enrollment rate will ↑

Variable	Coefficient
total contact count↑	0.43612
Self-initiated contact count↑	0.17495
Solicited contact count↓	-0.37087
Mail qualifying score (1 = very interested)↑	0.17161
Number of indicated extracurricular interest↑	0.67259
Time from first contact to enrollment date↓	-0.06979
5-year primary interest code rate↑	6.08524
5-year secondary interest code rate↑	6.36424
5-year enrollment rate from student's high school↑	11.84376
Campus visit code = 0↑	1.02076
Campus visit code = 1↑	2.30075
Attended campus recruitment event↑	0.99552
In-state↑	0.33732

Recommendation

- The admission team should focus on student's high school, the more students enroll in this university in the past five years, the more likely an admitted student will enroll this year
- The earlier students receive the offer letter, the less likely they will enroll in the university. However, the university can organize campus recruitment events and invite admitted students to visit the campus to attract new students
- The admission team should admit students with more extracurricular or they are in-state, because they will be more likely to enroll

Appendix - Method

- Logistic Regression
- Full Model→ Reduced Model
- Removed: stepAIC, p-value > 0.05
- Steps: Partition my data 70/30 (train / test split) → Recipe → Bake → Fit → Prep for Evaluation → Evaluate

.metric <chr>	.estimator <chr>	.estimate <dbl>	part <chr>
accuracy	binary	0.9640290	training
roc_auc	binary	0.9705458	training
accuracy	binary	0.9640825	testing
roc_auc	binary	0.9725269	testing

Appendix - Final Model

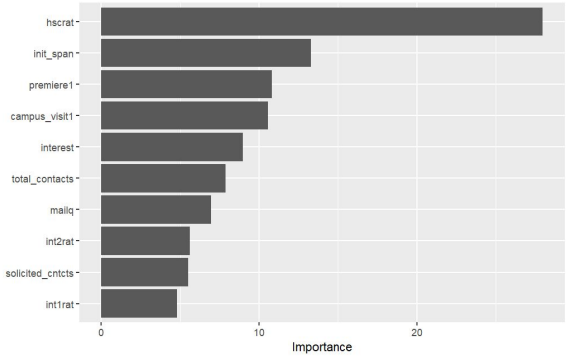
log(p/(1-p))= -6.2938 + 0.4361 (total_contacts) + 0.1750 (self_init_cntcts) - 0.3709 (solicited_cntcts) + 0.1716 (mailq) + 0.6726 (interest) - 0.0698 (init_span) + 6.0852 (int1rat) + 6.3642 (int2rat) + 11.8438 (hscrat) + 1.0208 (campus_visit1) + 2.3008 (campus_visit2) + 0.9955 (premiere1) + 0.3373 (instateY)

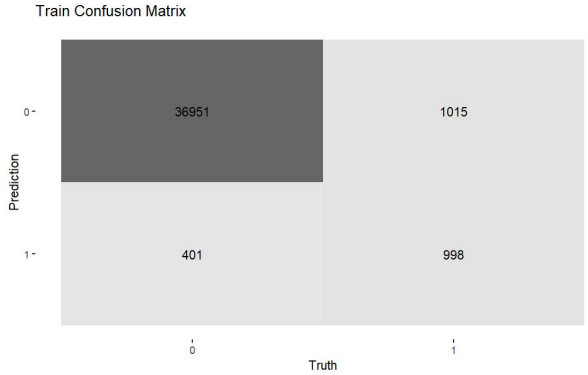
Appendix - Final Model

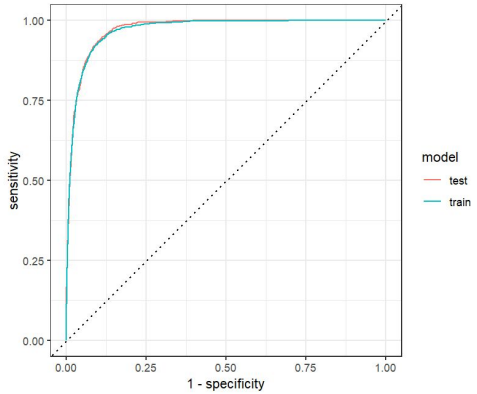
Coefficients:

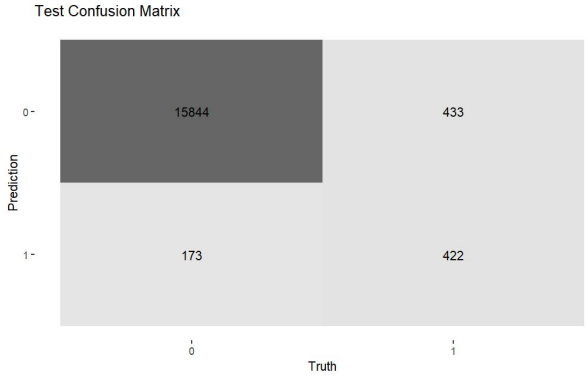
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-6.29381	0.17880	-35.201	< 2e-16	***
total_contacts	0.43612	0.05528	7.889	3.05e-15	***
self_init_cntcts	0.17495	0.05771	3.031	0.002435	**
solicited_cntcts	-0.37087	0.06741	-5.502	3.76e-08	***
mailq	0.17161	0.02466	6.959	3.44e-12	***
interest	0.67259	0.07500	8.967	< 2e-16	***
init_span	-0.06979	0.00525	-13.294	< 2e-16	***
int1rat	6.08524	1.26853	4.797	1.61e-06	***
int2rat	6.36424	1.13283	5.618	1.93e-08	***
hscrat	11.84376	0.42333	27.977	< 2e-16	***
campus_visit1	1.02076	0.09660	10.567	< 2e-16	***
campus_visit2	2.30075	0.61208	3.759	0.000171	***
premiere1	0.99552	0.09209	10.811	< 2e-16	***
instateY	0.33732	0.10491	3.215	0.001303	**

Appendix - Evaluation









Thank You