VANDERBILT UNIVERSITY

Cross Validation: Predicting College Completion

Kiss et al: Predicting Dropout Using High School and First-semester Academic Achievement Measures

- Measures of Model Fit: (p. 387)

- Accuracy

- Sensitivity (Precision)

- Specificity (Recall)

- AUC (.5= bad, 1=perfect)

Kiss et al: Predicting Dropout Using High School and First-semester Academic Achievement Measure

- 10 K-fold cross validation (p.388)

- Mean of cross-validated results reported

Kiss et al: Predicting Dropout Using High School and First-semester Academic Achievement Measures

TABLE II: Evaluation metrics using data available at the time of enrollment (S1), its supplementation with first-week features (S2) all available features (S3) and the first semester related features only (S4).

Scenario	ML model	Performance			
		Accuracy	Precision	Recall	AUC
(S1)	GBT	68.0%	67.0%	73.5%	0.729
	XGB	71.9%	71.6%	72.1%	0.787
	ANN	74.5%	73.0%	73.4%	0.815
(S2)	GBT	73.3%	72.0%	70.5%	0.789
	XGB	74.2%	72.6%	72.1%	0.816
	ANN	76.6%	75.9%	73.4%	0.823
(S3)	GBT	85.3%	85.5%	80.8%	0.920
	XGB	85.4%	86.1%	80.8%	0.920
	ANN	85.8%	86.3%	81.8%	0.916
(S4)	XGB	82.6%	82.0%	74.6%	0.892

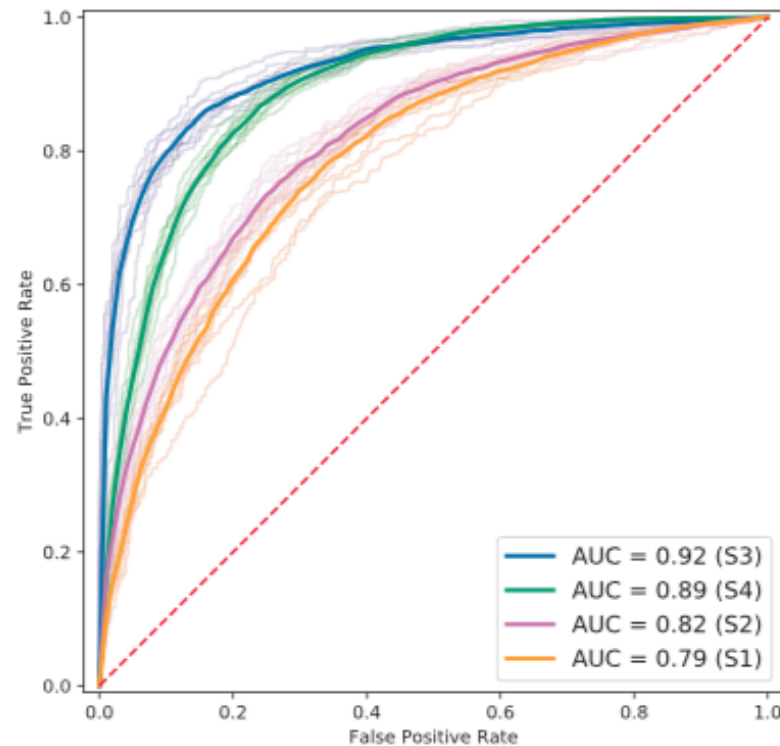# Kiss et al: Predicting Dropout Using High School and First-semester Academic Achievement Measures

Fig. 4: ROC curves of 10-fold cross-validation of XGB predictors corresponding to four modeling scenarios. The average ROC curves of the cross-validations are shown in darker shades.

Kiss et al: Predicting Dropout Using High School and First-semester Academic Achievement Measures