
Predicting Student Success

Rohan Doshi

Department of Computer Science
Princeton University
rkdoshi@princeton.edu

Kevin Liu

Department of Operations Research
and Financial Engineering
Princeton University
kevinliu@princeton.edu

Samhita Karnati

Department of Computer Science
Princeton University
skarnati@princeton.edu

Avinash Nayak

Department of Computer Science
Princeton University
anayak@princeton.edu

1 Introduction

2 Methods

2.1 Data Preprocessing and Missing Data Imputation

We downloaded the College Scorecard dataset from Kaggle as a CSV file that contains 124700 rows (representing universities) and 1700 columns (representing various attributes of the university).

2.2 Supervised Learning Methods

2.3 Unsupervised Learning Methods

2.4 Evaluation and validation

discuss r^2 and RMSE weigh merits of different evaluation metrics

To evaluate the performance of our regression models, we use r^2 to evaluate the correlation between predicted and true β values and RMSE to evaluate the relative magnitude of error.

In order to fine tune hyper-parameters for each regression model, we considered a set K of potential hyperparameter values, computed generalized error for each $k \in K$ using 3-fold cross validation, and then selected the k with the lowest generalization error.

3 Results

3.1 Preliminary Data Analysis

base on technical report furthermore data analysis

3.2 Supervised Learning: Regression Analysis

Feature Selection

each model makes assumptions about data, best regression models suggest coresponding assumptions are the most accurate.

054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107

3.3 Supervised Learning: Feature Importance

3.4 Unsupervised Learning: PCA and Clustering

get lower dimensional representation of data, reveal latent structures... cluster... find schools that are outliers (does this imply anything about income)... didn't fit on y...

4 Summary and Conclusion

References

- [1] Kaggle. US Dept of Education: College Scoreboard. <https://www.kaggle.com/kaggle/college-scorecard>. Accessed: 2016-04-17.