Uni-X

Regression Analysis on College Scorecard Data

Group members: Ahlam Hakami
Bin Luo
Qi Zhang

Outline

- Introduction
- Project Goals
- Data Descriptive Analysis
- Principle Component Analysis (PCA)
- Regression Analysis
- Conclusion
- References

Introduction

Why is repayment rate important?

- In the recent reports, there is 44.2 million Americans with student loan debt. [1]

- The federal government issues $1.45 trillion in total of student loan debt each year, and more than 90% of student debt today is in the form of federal loans.[1]

Data set

The data set [2] cover a wide range of elements for more than 75,000 institutions including research universities, state colleges and universities, private religious and liberal arts colleges, community and technical colleges, and others. These data are provided through federal reporting from institutions, data on federal financial aid, and tax information.

College Scorecard data (https://collegescorecard.ed.gov/data/)

Project Goals

1. What variables can affect repayment rate?

2. Estimate repayment rate based on other variables.

Data Descriptive Analysis

- General information
  - Data cover nearly 7 years (2007-2014)
  - More than 1,000, categorized into 9 groups

Variables related to our Data

TABLE I
NUMBER OF VARIABLES IN EACH CATEGORY

Category	Number of Variables
Repayment	131
School	170
Academics	228
Admission	25
Cost	65
Student	96
Completion	1031
Aid	40
Earnings	73

Data Descriptive Analysis

- Study variable (repayment rate)

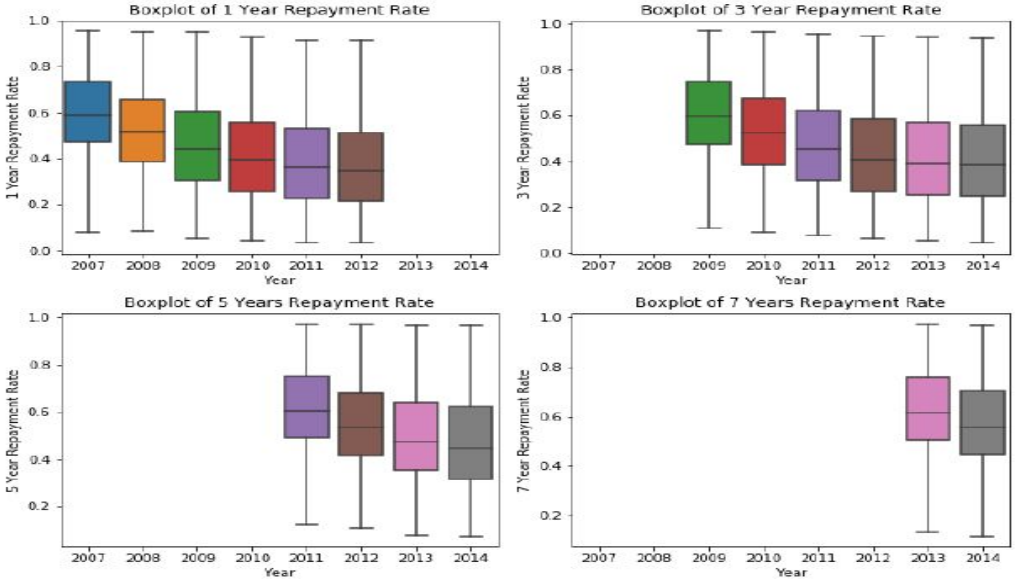
  - 1 year repayment rate
  - 3 years repayment rate
  - 5 years repayment rate
  - 7 years repayment rate

  - 1 year female repayment rate
  - 3 years female repayment rate
  - 5 years female repayment rate
  - 7 years female repayment rate
  - 1 year male repayment rate
  - 3 years male repayment rate
  - 5 years male repayment rate
  - 7 years male repayment rate

Data Descriptive Analysis

- Compare repayment rates in different years

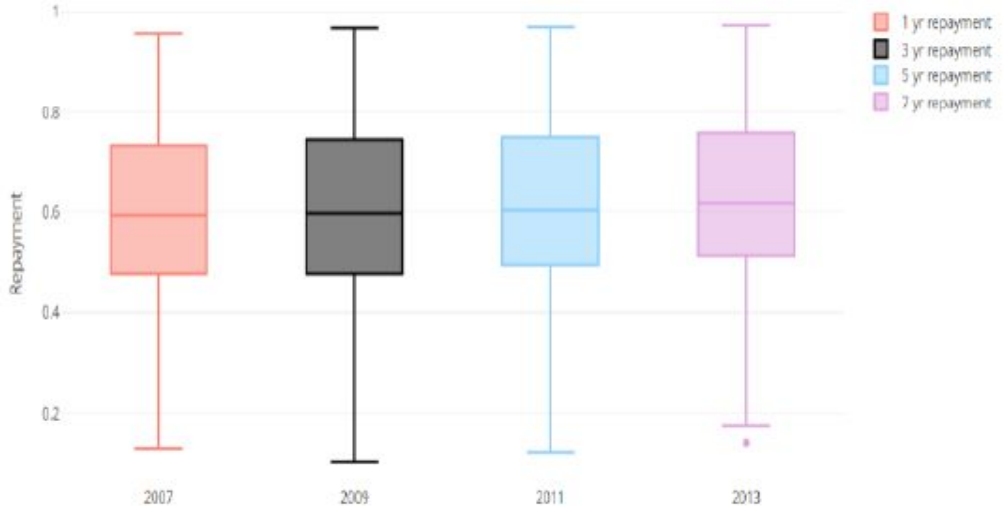
Fig. 1. REPAYMENT RATES IN DIFFERENT YEARS



Data Descriptive Analysis

- Compare different time periods repayment rate for the same group of graduates

Fig. 2. DIFFERENT TIME PERIODS REPAYMENT RATE FOR STUDENT WHO GRADUATED IN 2006



Data Descriptive Analysis

- Compare repayment rate means of male and female groups

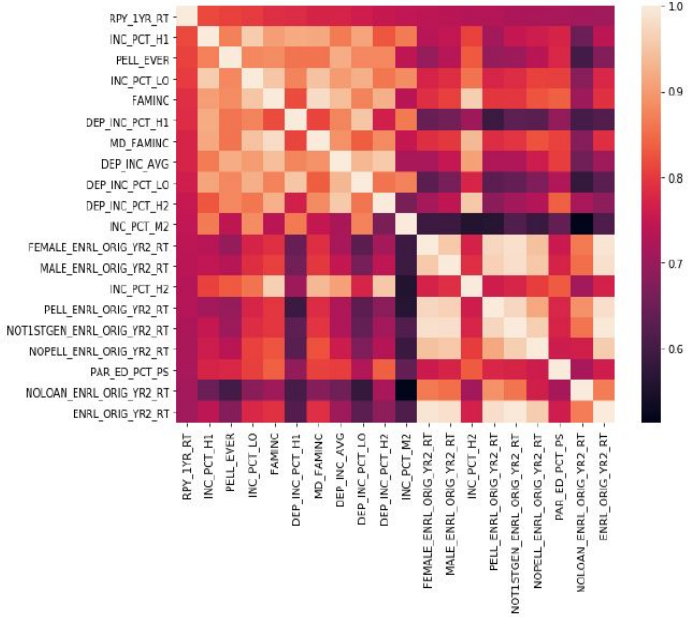
TABLE II
HYPOTHESIS TESTING FOR MEAN OF MALE AND FEMALE GROUPS

	Mean		H_0: $\mu_f=\mu_m$ H_a: $\mu_f \neq \mu_m$	
	Male	Female	T value	P value
1 Year	0.4525	0.4567	-2.4751	0.0133
3 Years	0.4759	0.4784	-1.5119	0.1306
5 Years	0.5335	0.5343	-0.3821	0.7024
7 Years	0.5987	0.5986	0.0588	0.9531

Data Descriptive Analysis

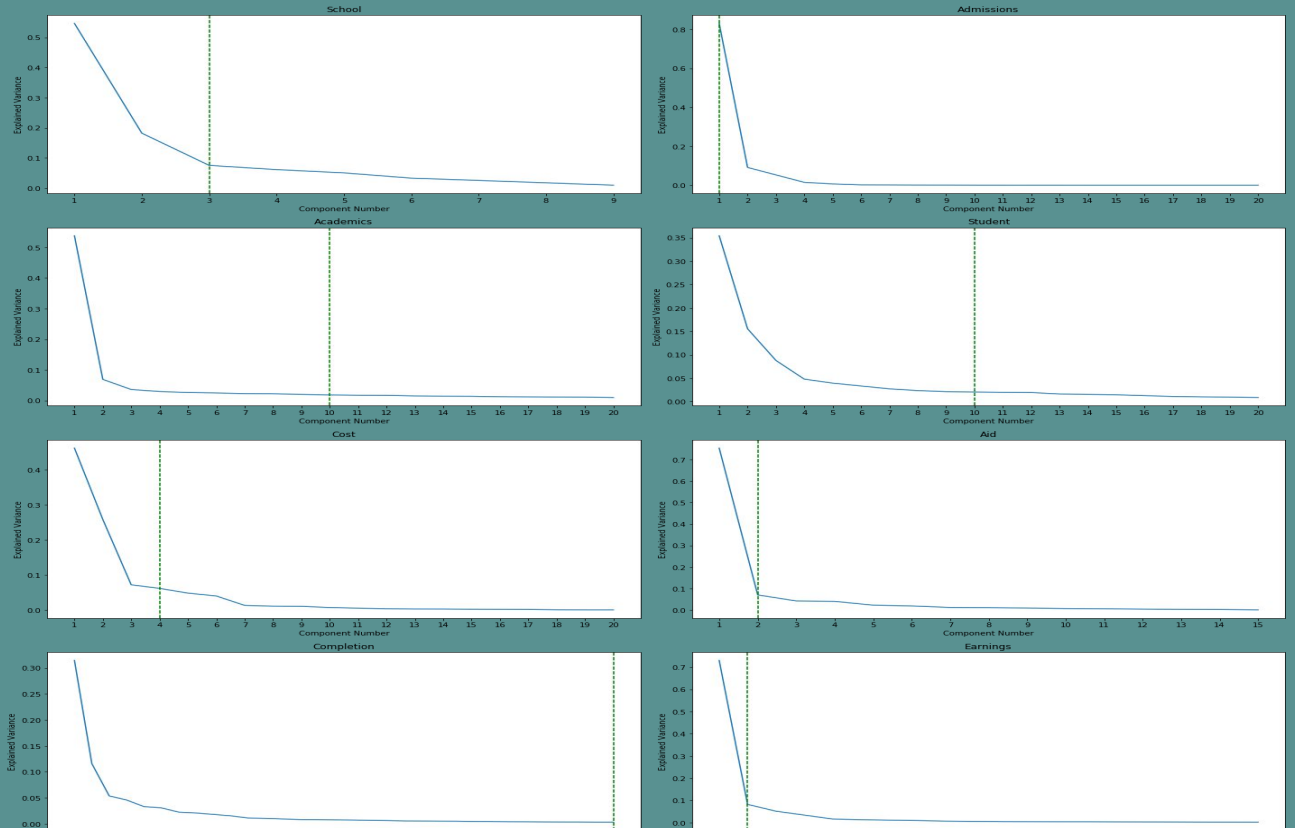
- Potential predictors

Fig. 3. TOP 20 HIGHTLY CORRELATED VARIABLES



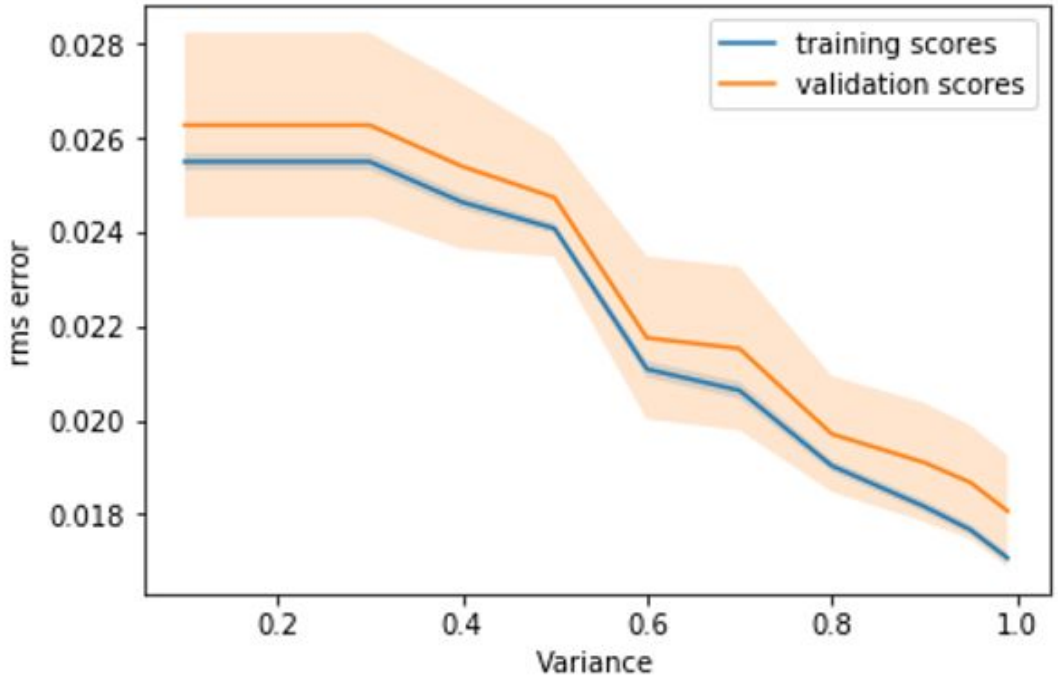
	desc	category	name
678	Percent still enrolled at original institution within 2 years	completion	ENRL_ORIG_YR2_RT
756	Percent of female students who were still enrolled at original institution within 2 years	completion	FEMALE_ENRL_ORIG_YR2_RT
769	Percent of male students who were still enrolled at original institution within 2 years	completion	MALE_ENRL_ORIG_YR2_RT
782	Percent of students who received a Pell Grant at the institution and who were still enrolled at ...	completion	PELL_ENRL_ORIG_YR2_RT
795	Percent of students who never received a Pell Grant at the institution and who were still enroll...	completion	NOPELL_ENRL_ORIG_YR2_RT
821	Percent of students who never received a federal loan at the institution and who were still enro...	completion	NOLOAN_ENRL_ORIG_YR2_RT
847	Percent of not-first-generation students who were still enrolled at original institution within ...	completion	NOT1STGEN_ENRL_ORIG_YR2_RT
1637	Percentage of aided students whose family income is between 0–30,000	student	INC_PCT_LO
1640	Percentage of students who are financially dependent and have family incomes between $0-30,000	student	DEP_INC_PCT_LO
1643	Aided students with family incomes between 48,001–75,000 in nominal dollars	student	INC_PCT_M2
1644	Aided students with family incomes between 75,001–110,000 in nominal dollars	student	INC_PCT_H1
1645	Aided students with family incomes between $110,001+ in nominal dollars	student	INC_PCT_H2
1648	Dependent students with family incomes between 75,001–110,000 in nominal dollars	student	DEP_INC_PCT_H1
1649	Dependent students with family incomes between $110,001+ in nominal dollars	student	DEP_INC_PCT_H2
1656	Percent of students whose parents' highest educational level was is some form of postsecondary e...	student	PAR_ED_PCT_PS
1661	Average family income of dependent students in real 2015 dollars.	student	DEP_INC_AVG
1835	Share of students who received a Pell Grant while in school	student	PELL_EVER
1844	Average family income in real 2015 dollars	student	FAMINC
1845	Median family income in real 2015 dollars	student	MD_FAMINC

Principle Component Analysis



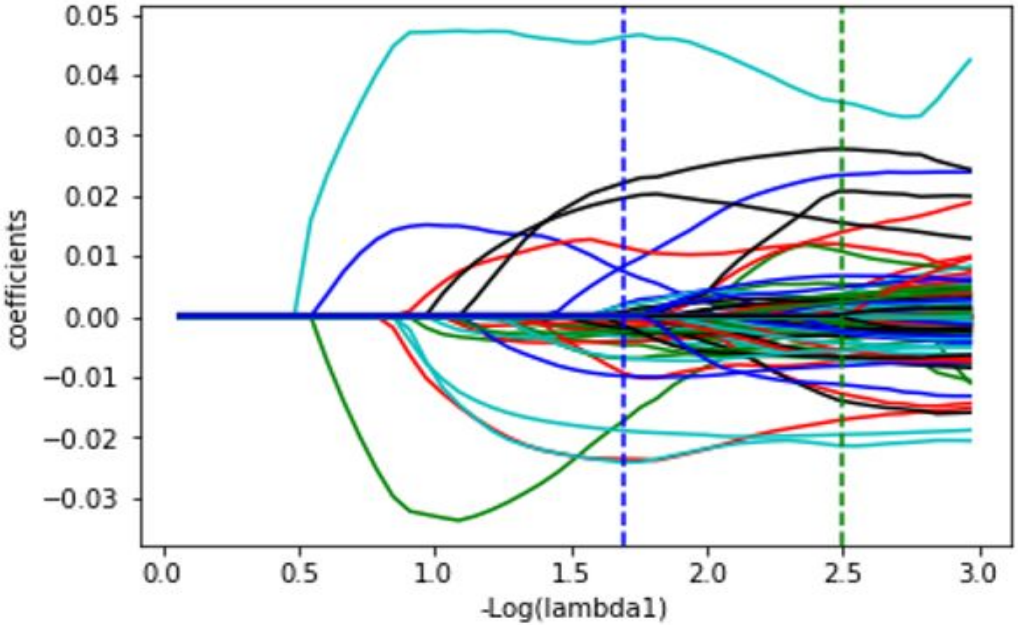
Linear Regression Analysis

Fig. 5. Display of Validation Curve with RMSE vs. Variance.



Elastic Net Regression

Fig. 6. Elastic Net Solution Path with Coefficients vs. $-log(\lambda_1)$. The green dash line represents the optimal tuning parameter and solution chosen by 5-fold cross validation; The blue dash line represents the parameter we manually tune.



Display of the Important Features Chosen by Elastic Net

	desc	category	name
25	Control of institution	school	CONTROL
531	Total share of enrollment of undergraduate degree-seeking students who are black	student	UGDS_BLACK
539	Total share of enrollment of undergraduate degree-seeking students who are white non-Hispanic	student	UGDS_WHITENH
624	Percentage of undergraduates who receive a Pell Grant	aid	PCTPELL
675	Percent withdrawn from original institution within 2 years	completion	WDRAW_ORIG_YR2_RT
677	Percent who transferred to a 2-year institution and withdrew within 2 years	completion	WDRAW_2YR_TRANS_YR2_RT
818	Percent of students who never received a federal loan at the institution and withdrew from origi...	completion	NOLOAN_WDRAW_ORIG_YR2_RT
845	Percent of not-first-generation students who transferred to a 4-year institution and withdrew wi...	completion	NOT1STGEN_WDRAW_4YR_TRANS_YR2_RT
846	Percent of not-first-generation students who transferred to a 2-year institution and withdrew wi...	completion	NOT1STGEN_WDRAW_2YR_TRANS_YR2_RT
858	Percent who transferred to a 4-year institution and withdrew within 3 years	completion	WDRAW_4YR_TRANS_YR3_RT
859	Percent who transferred to a 2-year institution and withdrew within 3 years	completion	WDRAW_2YR_TRANS_YR3_RT
1000	Percent of students who never received a federal loan at the institution and withdrew from origi...	completion	NOLOAN_WDRAW_ORIG_YR3_RT
1041	Percent who transferred to a 2-year institution and withdrew within 4 years	completion	WDRAW_2YR_TRANS_YR4_RT
1182	Percent of students who never received a federal loan at the institution and withdrew from origi...	completion	NOLOAN_WDRAW_ORIG_YR4_RT
1637	Percentage of aided students whose family income is between $0-30,000	student	INC_PCT_LO
1640	Percentage of students who are financially dependent and have family incomes between $0-30,000	student	DEP_INC_PCT_LO
1641	Percentage first-generation students	student	PAR_ED_PCT_1STGEN
1643	Aided students with family incomes between $48,001-75,000 in nominal dollars	student	INC_PCT_M2
1644	Aided students with family incomes between $75,001-110,000 in nominal dollars	student	INC_PCT_H1
1655	Percent of students whose parents' highest educational level is high school	student	PAR_ED_PCT_HS
1661	Average family income of dependent students in real 2015 dollars.	student	DEP_INC_AVG
1835	Share of students who received a Pell Grant while in school	student	PELL_EVER
1843	Share of first-generation students	student	FIRST_GEN
1844	Average family income in real 2015 dollars	student	FAMINC
1845	Median family income in real 2015 dollars	student	MD_FAMINC
1895	Median earnings of students working and not enrolled 6 years after entry	earnings	MD_EARN_WNE_P6

Random Forest Regression

TABLE III
RMSE FROM PCA-LS, ELASTICNET AND RANDOMFOREST COMPUTED BY CROSS VALIDATION

Method	PCA-LS	Elastic Net	Random Forest
RMSE	0.0197	0.0194	0.0170

Display of the Important Features Chosen by Random Forest

	desc	category	name
2	6-digit OPE ID for institution	root	OPEID6
531	Total share of enrollment of undergraduate degree-seeking students who are black	student	UGDS_BLACK
540	Total share of enrollment of undergraduate degree-seeking students who are black non-Hispanic	student	UGDS_BLACKNH
624	Percentage of undergraduates who receive a Pell Grant	aid	PCTPELL
675	Percent withdrawn from original institution within 2 years	completion	WDRAW_ORIG_YR2_RT
677	Percent who transferred to a 2-year institution and withdrew within 2 years	completion	WDRAW_2YR_TRANS_YR2_RT
688	Percent of low-income (less than $30,000 in nominal family income) students withdrawn from origi...	completion	LO_INC_WDRAW_ORIG_YR2_RT
831	Percent of first-generation students withdrawn from original institution within 2 years	completion	FIRSTGEN_WDRAW_ORIG_YR2_RT
844	Percent of not-first-generation students withdrawn from original institution within 2 years	completion	NOT1STGEN_WDRAW_ORIG_YR2_RT
857	Percent withdrawn from original institution within 3 years	completion	WDRAW_ORIG_YR3_RT
859	Percent who transferred to a 2-year institution and withdrew within 3 years	completion	WDRAW_2YR_TRANS_YR3_RT
948	Percent of male students withdrawn from original institution within 3 years	completion	MALE_WDRAW_ORIG_YR3_RT
1637	Percentage of aided students whose family income is between $0–30,000	student	INC_PCT_LO
1640	Percentage of students who are financially dependent and have family incomes between $0-30,000	student	DEP_INC_PCT_LO
1643	Aided students with family incomes between $48,001–75,000 in nominal dollars	student	INC_PCT_M2
1656	Percent of students whose parents' highest educational level was is some form of postsecondary e...	student	PAR_ED_PCT_PS
1657	Number of applications is greater than or equal to 2	student	APPL_SCH_PCT_GE2
1661	Average family income of dependent students in real 2015 dollars.	student	DEP_INC_AVG
1743	The median debt for female students	aid	FEMALE_DEBT_MDN
1745	The median debt for first-generation students	aid	FIRSTGEN_DEBT_MDN
1764	Cumulative loan debt at the 75th percentile	aid	CUML_DEBT_P75
1835	Share of students who received a Pell Grant while in school	student	PELL_EVER
1836	Average age of entry	student	AGE_ENTRY
1840	Share of married students	student	MARRIED
1844	Average family income in real 2015 dollars	student	FAMINC
1845	Median family income in real 2015 dollars	student	MD_FAMINC

Conclusion

- Student family income, student withdrawing and enrolling rate are important factor to affect the repayment rate.

- All three regression methods: PCA-LS, Elastic Net, Random Forest, are able to predict the repayment rate with standard error < 2%.

References

- Federal Reserve System (2017 Nov)
  Available: https://www.federalreserve.gov/releases/g19/current/default.htm

- The Department of Education. (2017 Sep.) Collage Scorecard Data.
  Available: https://collegescorecard.ed.gov/data/