Predicting student loan credit recovery using machine learning methodologies

Kim (2024)

Presenter
Hyeongjun Kim, Assistant Professor
Department of Business Administration, Yeungnam University
hkim@yu.ac.kr
053-810-2740

Abstract

☐ Main Findings

- This study forecasts failure of student loan credit recovery program participants using machine learning algorithms with account-level dataset
  - Loan defaulters are not for conventional credit assessments
  - However, credit recovery programs need to predict participant's ability to repay

- We find that the artificial neural networks algorithm performs best in predicting credit risks of student loan loan defaulters
  - Predicted failure risks match the actual failure rates of credit recovery
  - Logistic regression can be an efficient alternative considering the computational resources

- However, the information currently used in credit recovery programs may not be sufficient for assessing bad debtors' ability to recover their credit

- Our findings provide valuable implications for policymakers and financial institutions who seek to improve student loan credit recovery rates

Motivation

☐ Student loans are very hot topic in the U.S. nowadays

- The United States announced COVID-19 relief from the Federal Student Aid and suspended student loan repayment from March 2020 after the last pandemic.
  - The size of student loans in the U.S. increased from $0.96 trillion in 2011 to $1.76 trillion in 2022
  - These measures are evaluated not only to ease the repayment burden of students but also to prevent the insolvency of financial institutions.

- In August 2022, Biden administration announced student loan relief
  - President Biden pledged to write off student loans in the 2020 presidential election.
  - It allowed students who received Pell Grant scholarships with an annual income of less than $125,000 to write off up to $20,000 and others up to $10,000 of their student loan principal.
  - However, in June 2023, the U.S. Supreme Court ruled 6-3 against the administration and put the brakes on it, saying that congressional approval should come first.

- In July 2023, Biden administration begins student loan debt forgiveness
  - The U.S. Department of Education announced that it would write off $39 billion in debt for 804,000 people who had received federal student loans and repaid them for more than 20-25 years.

Motivation

☐ Why should we pay attention to student loan credit recovery?

- Total student loan debt and average student loan debt increases in U.S.
  - In Korea, total student loan debt decreased overall with the expansion of national grants, but living expenses loans decreased less than tuition loans

- Stiglitz pointed out that "the crisis that is about to break out involves student debt and how we finance higher education"
  - Stiglitz, J. E., "Student Debt and the Crushing of the American Dream," New York Times (2013.05.12)

- Unlike personal loans, we should not identify which students are likely to default so that they could be declared ineligible for student loans (Gross et al., 2009)
  - There is a greater risk of default in providing loans to low- and moderate-income students who often come from families with weak credit histories
  - Even so, prohibiting student loans for students at high risk of default undermines the very original purpose of the student loan program: expanding educational opportunities

- Researches on student loan credit recovery using large data sets and rigorous statistical methods are needed

Motivation

☐ Student Loan Credit Recovery Program in Various Countries

- In the U.S. – Student Loan Rehabilitation
  - Allows to recover credit score and does not add additional costs to borrower's principal
  - Generally, when designing a repayment plan, the borrower plan to repay 10% of his/her discretionary income (disposable income - basic living expenses) every month

- In Canada – the Repayment Assistance Plan (RAP)
  - Repayment begins after a six-month non-payment period (depends on the type of loan) following the end of the borrower's studies. At this time, the borrower can adjust his/her repayment amount by adjusting a few repayment options, such as interest rate, repayment method, and payment date.
  - If the borrower have difficulty repaying, he/she can get help through the Repayment Assistance Plan (RAP) or Repayment Assistance Plan for Borrowers with Disabilities (RAP-D) program.

- Other Countries
  - In the UK and Australia, student loan repayment is linked to income tax, so repayment is exempt if the borrower has no income.
  - In Germany and Sweden, public universities do not need tuition fees (but they may charge other fees such as administration or student union fees)

Motivation

☐ Student Loan Credit Recovery Program in Korea

- If a student loan borrower does not repay his/her interest and principal payments more than 6 months in a row, the borrower will become a long-term delinquent
  - In that case, this borrower should repay his/her remaining loan balances immediately regardless of its original maturity, or the student loan will be in default

- Long-term delinquents of student loans can use student loan credit recovery program, which allows them to repay their defaulted loans in monthly instalments
  - Repayment period should be less than 10 years
  - If loan size is larger than 20 million won (about 15,000 USD), repayment period should be less than 20 years
  - Initial payment should be at least 2 percent of the loan size, and about ten percent is recommended

- If a delinquent borrower use student loan credit recovery program
  - The borrower's relevant credit information will be removed
  - The borrower can repay his/her delinquent loan with a longer repayment period

Literature Review

☐ Related Studies on Credit Risks of Student Loans

- Gross et al. (2009)
  - Nearly all studies that considered the age of the student concluded that as age increases, so does the likelihood of loan default, even after controlling for other important factors such as income
  - Whatever the type of institution, the more a student borrows the greater the chance of default
  - The majority of research suggested that completing a postsecondary program is the strongest single predictor of not defaulting regardless of institutional type
- Mueller and Yannelis (2019)
  - Shifts in the composition of student loan borrowers and the massive collapse in home prices during the Great Recession can each account for approximately 30% of the rise in student loan defaults.
- Looney and Yannelis (2019)
  - Increases in credit limits and expansions in credit availability resulted in rising borrowing amounts, and that the share of borrowers holding very large balances has surged
- Han et al. (2015)
  - Default of student loans is a function of gender, major, loan balance, etc., and new variables such as grace period and repayment period also affect student loan default in Korea

Data

☐ Micro Dataset on Student Loan Credit Recovery

- We analyze 171,899 student loans that applied for the student loan credit recovery program from August 2006 to May 2019
  - All accounts have been monitored for over 24 months, until May 2022
  - Student loans and their credit recovery program are managed by Korea Student Aid Foundation

- Our dataset contains following information:
  - Demographic Variables: Gender, Age, Income Level, Multi-child, Disability, Region
  - Original Loan Characteristics: Delinquent Period, Income Contingent Loan Dummy, Any Legal Actions (e.g., execution, lawsuit, or sized)
  - Rehabilitation Characteristics: Loan Amount, Contract Period, Debt Relief Dummy, Initial Payment, Payment, Number of Settlement
  - Educational Characteristics: Public School Dummy, Institutional Dummies (e.g., four-year, two-year, graduate school, others, and no related information), Program of Study
  - Macroeconomic Variable: Unemployment Rate

Table 1. Descriptive Statistics

Variables	Mean	Std. Dev.	Min	25%	50%	75%	Max
Demographic Variables							
D_FEMALE	0.5000	0.5000	0.0000	0.0000	0.0000	1.0000	1.0000
AGE	27.2477	5.8166	17.0000	24.0000	26.0000	29.0000	67.0000
D_DISABLED	0.0021	0.0459	0.0000	0.0000	0.0000	0.0000	1.0000
D_CHILDREN	0.0500	0.2179	0.0000	0.0000	0.0000	0.0000	1.0000
D_REGION	0.5297	0.4991	0.0000	0.0000	1.0000	1.0000	1.0000
D_INCOME99	0.5403	0.4984	0.0000	0.0000	1.0000	1.0000	1.0000
D_INCOME00_03	0.2515	0.4339	0.0000	0.0000	0.0000	1.0000	1.0000
D_INCOME04_07	0.1218	0.3270	0.0000	0.0000	0.0000	0.0000	1.0000
D_INCOME08_10	0.0865	0.2812	0.0000	0.0000	0.0000	0.0000	1.0000
Original Loan Characteristics							
LOAN_AMOUNT	6.2320	5.9616	0.0000	2.6217	4.3771	7.8989	119.3319
DELINQUENT_PERIOD	19.9484	19.8312	0.0000	3.0000	14.0000	31.0000	146.0000
D_ICL	0.0046	0.0675	0.0000	0.0000	0.0000	0.0000	1.0000
D_ACTION	0.0656	0.2476	0.0000	0.0000	0.0000	0.0000	1.0000
Rehabilitation Program Characteristics							
D_DEBTRELIEF	0.4712	0.4992	0.0000	0.0000	0.0000	1.0000	1.0000
CONTRACT_PERIOD	56.5273	40.8288	0.0000	23.0000	48.0000	91.0000	312.0000
INITIAL_PMT	0.3843	0.7057	0.0000	0.1100	0.2000	0.4000	30.1757
PAYMENT	0.1718	0.5016	0.0000	0.0500	0.1000	0.1500	30.1757
N_SETTLEMENT	1.4516	0.7579	1.0000	1.0000	1.0000	2.0000	9.0000

Table 1. Descriptive Statistics

Variables	Mean	Std. Dev.	Min	25%	50%	75%	Max
Educational Characteristics							
D_PUBLIC	0.0942	0.2920	0.0000	0.0000	0.0000	0.0000	1.0000
D_SCHOOL0	0.4310	0.4952	0.0000	0.0000	0.0000	1.0000	1.0000
D_SCHOOL1	0.3348	0.4719	0.0000	0.0000	0.0000	1.0000	1.0000
D_SCHOOL2	0.1058	0.3076	0.0000	0.0000	0.0000	0.0000	1.0000
D_SCHOOL3	0.1249	0.3306	0.0000	0.0000	0.0000	0.0000	1.0000
D_SCHOOL4	0.0036	0.0596	0.0000	0.0000	0.0000	0.0000	1.0000
D_MAJOR0	0.2311	0.4215	0.0000	0.0000	0.0000	0.0000	1.0000
D_MAJOR1	0.2169	0.4121	0.0000	0.0000	0.0000	0.0000	1.0000
D_MAJOR2	0.1914	0.3934	0.0000	0.0000	0.0000	0.0000	1.0000
D_MAJOR3	0.1407	0.3477	0.0000	0.0000	0.0000	0.0000	1.0000
D_MAJOR4	0.1130	0.3166	0.0000	0.0000	0.0000	0.0000	1.0000
D_MAJOR5	0.0376	0.1903	0.0000	0.0000	0.0000	0.0000	1.0000
D_MAJOR6	0.0308	0.1727	0.0000	0.0000	0.0000	0.0000	1.0000
D_MAJOR7	0.0024	0.0491	0.0000	0.0000	0.0000	0.0000	1.0000
D_MAJOR8	0.0361	0.1866	0.0000	0.0000	0.0000	0.0000	1.0000
Macroeconomic Variables							
UNRATE	3.4792	0.3022	2.9000	3.2000	3.5000	3.7000	4.6000
Results							
D_FAILURE24M	0.2021	0.4016	0.0000	0.0000	0.0000	0.0000	1.0000

Table 1. Descriptive Statistics

Variables	Description	Mean	Std. Dev.
Demographic Variables			
D_FEMALE	equals 1 if gender of the borrower is female	0.5000	0.5000
AGE	Age of the borrower at the time of rehabilitation	27.2276	5.8127
D_DISABLED	equals 1 if the borrower is disabled	0.0021	0.0459
D_CHILDREN	equals 1 if the borrower has multi-child	0.0500	0.2179
D_REGION	equals 1 if the borrower leaves in the outside of Seoul metropolitan area	0.5297	0.4991
D_INCOME99	equals 1 if there is no income-related information	0.5403	0.4984
D_INCOME00_03	equals 1 if the borrower's income level is low	0.2515	0.4339
D_INCOME04_07	equals 1 if the borrower's income level is medium	0.1218	0.3270
D_INCOME08_10	equals 1 if the borrower's income level is high	0.0865	0.2812
Original Loan Characteristics			
LOAN_AMOUNT	Amount of the delinquent loan including accrued interests (in thousand won)	6.2320	5.9616
DELINQUENT_PERIOD	Number of months after the borrower's first delinquency	19.3968	19.8220
D_ICL	equals 1 if the original loan is an income contingent loan	0.0046	0.0675
D_ACTION	equals 1 if there were any legal actions before rehabilitation	0.0656	0.2476
Rehabilitation Program Characteristics			
D_DEBTRELIEF	equals 1 if there was any debt relief in the rehabilitation program	0.4712	0.4992
CONTRACT_PERIOD	Number of monthly payments until the maturity of rehabilitated loan	56.5273	40.8288
INITIAL_PMT	Amount of initial payment (in thousand won)	0.3843	0.7057
PAYMENT	Amount of monthly payment (in thousand won)	0.1718	0.5016
N_SETTLEMENT	Number of rehabilitation program settlement	1.4516	0.7579

Table 1. Descriptive Statistics – cont’d

Variables	Description	Mean	Std. Dev.
Educational Characteristics			
D_PUBLIC	equals 1 if the institution is a public school	0.0942	0.2920
D_SCHOOL0	equals 1 if the institution is a four-year postsecondary institution	0.4310	0.4952
D_SCHOOL1	equals 1 if the institution is a two-year postsecondary institution	0.3348	0.4719
D_SCHOOL2	equals 1 if the institution is a graduate school	0.1058	0.3076
D_SCHOOL3	equals 1 if the institution is another type of postsecondary institution	0.1249	0.3306
D_SCHOOL4	equals 1 if there is no institution-related information	0.0036	0.0596
D_MAJOR0	equals 1 if the program of study is sociology	0.2311	0.4215
D_MAJOR1	equals 1 if the program of study is engineering	0.2169	0.4121
D_MAJOR2	equals 1 if the program of study is arts and physical education	0.1914	0.3934
D_MAJOR3	equals 1 if the program of study is humanities	0.1407	0.3477
D_MAJOR4	equals 1 if the program of study is science	0.1130	0.3166
D_MAJOR5	equals 1 if the program of study is education	0.0376	0.1903
D_MAJOR6	equals 1 if the program of study is medicine and pharmacy	0.0308	0.1727
D_MAJOR7	equals 1 if the program of study is not in the previous areas	0.0024	0.0491
D_MAJOR8	equals 1 if there is no major-related information	0.0361	0.1866
Results			
D_FAILURE24M	equals 1 if the credit recovery program is failed in 24 months	0.2046	0.4034

Data

☐ Data

- In this study, we predict whether applicants for the student loan credit recovery program will fail to repay on a predetermined payment schedule within 24 months
  - Failure rates increase sharply in the first few months and then decline
  - After 24 months, the failure rate stabilizes in the long-run

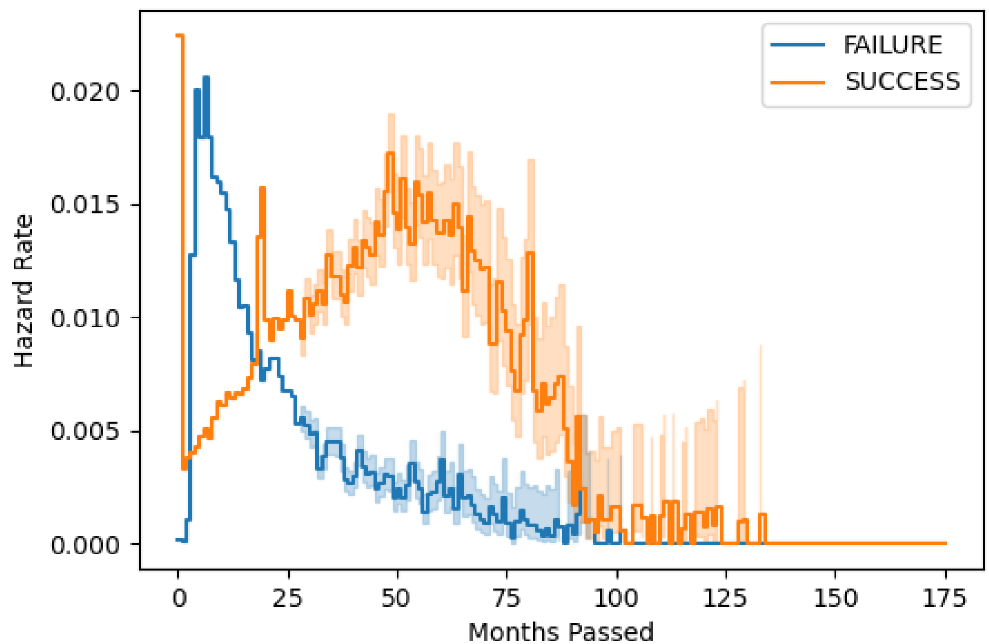


FIGURE 2. Monthly failure rates of the credit recovery program over time

Methodology

☐ Cox Proportional Hazard

- The survival function of T is expressed as $S(t) = \Pr(T \geq t)$, and the hazard function $\lambda_j(t)$, which specifies the instantaneous failure at t attributable to the j^{th} outcome, is written as

$$\lambda_j(t) = \lim_{\Delta t \to 0^+} \frac{\Pr(t \leq T < t + \Delta t, J = j | T \geq t)}{\Delta t}$$

- Assuming that there are no simultaneous terminations attributable to more than one cause, the overall instantaneous termination probability is $\lambda(t) = \Sigma_j \lambda_j(t)$.

- The hazard function of a type j failure in the Cox proportional hazard model takes the form $\lambda_j(t|X_{it}) = \lambda_{j0}(t)e^{X_{it}\beta_j}$ where $\lambda_{j0}(t)$ is the baseline hazard function, X_{it} is the explanatory variable for the i^{th} observation at time t, and β_j is the vector of unknown regression parameters.

Methodology

☐ Logistic Regression

- It can produce a probability score of default
- The probability of the observation $n = 1, \dots, N$ declaring default, P_n, takes the following form:

$$P_n(y_n = 1) = 1 / (1 + e^{-z})$$
$$= 1 / \{1 + \exp[- (\beta_0 + \beta_1 X_{1,n} + \beta_2 X_{2,n} + \dots + \beta_m X_{m,n})]\}$$

where

y_n = 1 if entity n has defaulted and 0 if entity n has not defaulted,
$P_n(y_n = 1)$ = probability of failure for entity n,
$\beta_1, \beta_2, \dots, \beta_m$ = slope coefficients,
$X_{1,n}, X_{2,n}, \dots, X_{m,n}$ = explanatory variables for entity n.

- Thus, the likelihood function is given by $L = \prod_{n=1}^{N} F(\beta' X_n)^{y_n} (1 - F(\beta' X_n))^{1-y_n}$
where

$$F(\beta' X_n) = 1 / \{1 + \exp[- (\beta_0 + \beta_1 X_{1,n} + \beta_2 X_{2,n} + \dots + \beta_m X_{m,n})]\}$$

- The maximum likelihood technique is used to estimate the coefficients.

Methodology

☐ Support Vector Machine (SVM)

- Classification using a separating *hyperplane*
  - Hyperplane: in a *p*-dimensional space, a *hyperplane* is a flat affine subspace of dimension *p*-1 (James et al., 2013)
- SVM with linear kernel:

$$\max_{\beta_0,\beta_1,\ldots,\beta_p,\epsilon_1,\ldots,\epsilon_n} M$$

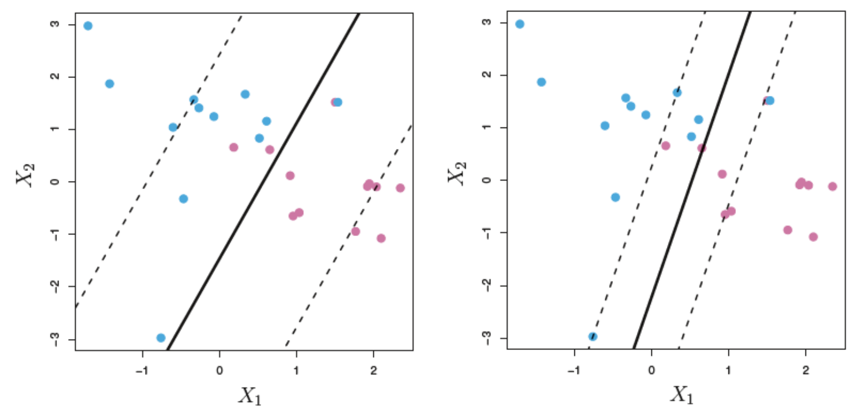
$$\text{subject to} \quad \Sigma_{j=1}^{p}\beta_j^2 = 1$$

$$y_i\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}\right) \geq M\left(1-\epsilon_i\right)$$

$$\epsilon_i \geq 0, \quad \Sigma_{i=1}^{n}\epsilon_i \leq C$$

where C is a nonnegative tuning parameter

FIGURE. Support vector classifier with different tuning parameter *C*
Notes: This figure shows examples of support vector machine with different tuning parameter *C*. (Source: James et al. (2013))

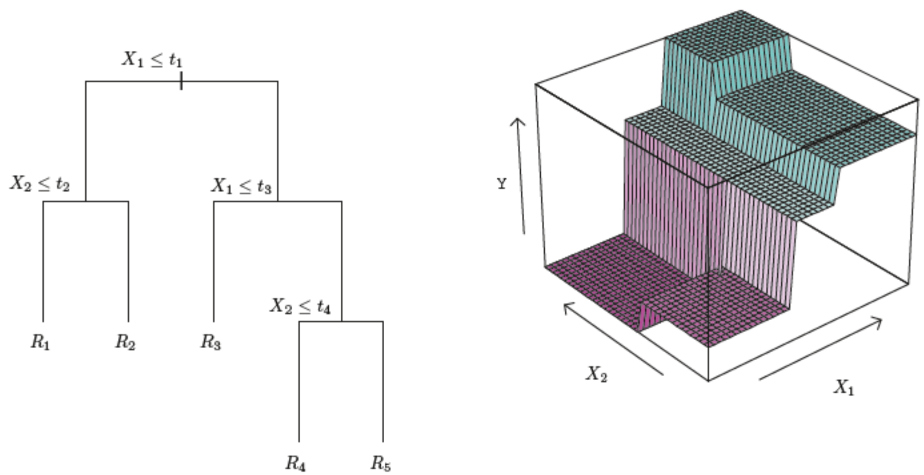


Methodology

☐ Decision Trees and Random Forest

- Solve the problem by using tree representation
  - Divide the predictor space - the set of possible values for $X_1, X_2, \ldots, X_p$ - into J distinct and non-overlapping regions $R_1, R_2, \ldots, R_J$
  - For every observation that falls into the region R_J, we make the same prediction

FIGURE. Decision Trees with Two-dimensional Feature Space
Notes: This figure shows examples of decision trees with two-dimensional feature space. The left shows a tree model corresponding to the recursive binary splitting, and the right shows a perspective plot of the prediction corresponding to that tree. (Source: James et al. (2013))



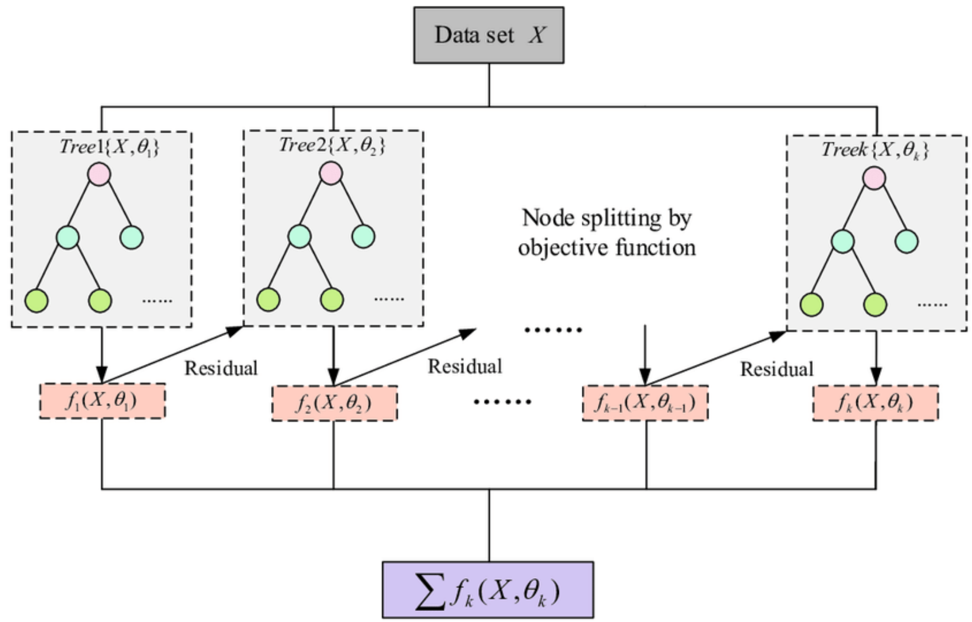
Methodology

☐ XGBoost (Extreme Gradient Boosting)

- An advanced implementation of the gradient boosting algorithm, designed for speed and performance
  - Optimized for both computation speed and model accuracy
  - Can handle large-scale datasets efficiently
- Includes L1 (Lasso) and L2 (Ridge) regularization techniques to prevent overfitting

FIGURE. Flow chart of XGBoost

Notes: This figure shows depicts the overall process of building a model using XGBoost algorithm. It starts with data input, followed by data preprocessing which includes handling missing values and feature transformation. The next step involves the initialization of the first model. The algorithm then iteratively adds new trees, optimizing the objective function using gradient descent. Each iteration focuses on correcting the errors from the previous model. The process continues until the stopping criteria are met, resulting in a final ensemble model. (Source: Guo et al., 2020)



Methodology

☐ Artificial Neural Networks (ANN)

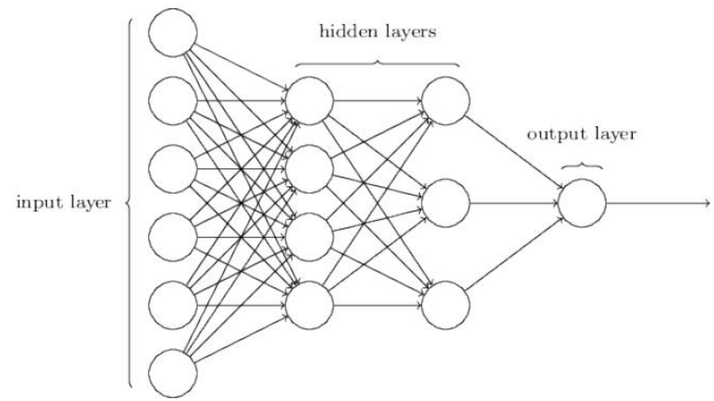
- Consists of the following artificial neurons (or *perceptrons*):

$$y_i = output = \begin{cases} 0 \quad if \ \sum_j w_j x_j \leq threshold \\ 1 \quad if \ \sum_j w_j x_j > threshold \end{cases}$$

- Can solve regression analysis as well as classification problems by linking perceptrons and finding optimal weights

- Rapid development after the seminal work of Hinton, Osindero, and Teh (2006).

FIGURE. Artificial neural network with two hidden layers
Notes: This figure shows a conceptual image of artificial neural network, four-layer network with two hidden layers.
(Source: Nielsen (2015))



Empirical Results

☐ Forecasting Performances

- Table 5 presents the out-of-sample forecasting performances using various classification algorithms
- We are dealing with an imbalanced dataset!
  - Test dataset consists of 34,380 observations and contains 7,117 failures, 20.7% of the total
  - We need to focus on the *area under the receiver operating characteristic curve*, or AUC
- Artificial Neural Networks may be the most suitable algorithm when we predict a personal credit recovery will fail within 24 months
- But no model can make better predictions than no information rate of 79.3%
  - bad debtor's features currently used in credit recovery programs may not be sufficient for assessing their ability to complete a credit recovery program.

TABLE 5. Forecasting performances by methodology

Model	Accuracy	Precision	Recall	F	AUC
CPH	0.7934	0.3200	0.0114	0.0220	0.5026
LR	0.5790	0.2941	0.7547	0.4232	0.6442
RF	0.6732	0.3305	0.5819	0.4215	0.6393
XGB	0.7734	0.3737	0.1588	0.2228	0.5451
NN	0.5533	0.2902	0.8181	0.4284	0.6516

Empirical Results

☐ Importance Rankings - RF and XGBoost

- Table 6 shows the rankings of Top 20 numeric importance measures and scores of explanatory variables used in each model
- Importance measures are quite similar for those in top 10 variables
- Unemployment rate is very highly ranked

TABLE 6. Importance rankings of the explanatory variables used in machine learning algorithms (Top 20 variables)

	RF		XGB	
Rank 1	N_SETTLEMENT	(0.30)	N_SETTLEMENT	(0.47)
Rank 2	DELINQUENT_PERIOD	(0.19)	PAYMENT	(0.05)
Rank 3	UNRATE	(0.11)	UNRATE	(0.05)
Rank 4	PAYMENT	(0.08)	D_DEBTRELIEF	(0.04)
Rank 5	D_DEBTRELIEF	(0.07)	DELINQUENT_PERIOD	(0.03)
Rank 6	INITIAL_PMT	(0.06)	AGE	(0.03)
Rank 7	CONTRACT_PERIOD	(0.06)	D_ICL	(0.02)
Rank 8	AGE	(0.05)	D_MAJOR8	(0.02)
Rank 9	LOAN_AMOUNT	(0.03)	D_MAJOR6	(0.02)
Rank 10	D_FEMALE	(0.01)	INITIAL_PMT	(0.01)
Rank 11	D_REGION	(0.01)	D_CHILDREN	(0.01)
Rank 12	D_INCOME99	(0.00)	D_ACTION	(0.01)
Rank 13	D_ACTION	(0.00)	D_MAJOR5	(0.01)
Rank 14	D_SCHOOL1	(0.00)	D_INCOME08_10	(0.01)
Rank 15	D_INCOME00_03	(0.00)	D_MAJOR4	(0.01)
Rank 16	D_SCHOOL0	(0.00)	D_FEMALE	(0.01)
Rank 17	D_MAJOR1	(0.00)	D_INCOME04_07	(0.01)
Rank 18	D_SCHOOL3	(0.00)	D_PUBLIC	(0.01)
Rank 19	D_MAJOR8	(0.00)	CONTRACT_PERIOD	(0.01)
Rank 20	D_MAJOR0	(0.00)	D_SCHOOL3	(0.01)

Results

☐ Comparing with Original Loan Defaults

- Gender
  - Original: men are more likely than women to default on their original loans (Flint, 1997; Podgursky et al., 2002; Woo, 2002a, 2002b)
  - Credit recovery: men are also more likely than women to default on their rehabilitated loans and less likely to repay their rehabilitated loans

- Age
  - Original: as age increases, so does the likelihood of original loan default, even after controlling for other important factors such as income (Christman, 2000; Flint, 1997; Harrast, 2004; Herr & Burt, 2005; Podgursky et al., 2002; Steiner & Teszler, 2005; Woo, 2002a, 2002b).
  - Credit recovery: as borrower's age increases, the likelihood of rehabilitated loan default decreases and the likelihood of successful repay increases

Results

☐ Comparing with Original Loan Defaults

- Income
  - Original: the higher the family income the lower the likelihood the student will default (Knapp & Seaks, 1992; Wilms et al., 1987; Woo, 2002a, 2002b)
  - Credit recovery:
    - borrowers with income information are more likely to default and repay their rehabilitated loans
    - the likelihood of default decreases as income increases
    - the likelihood of repay increases as income increases

- Debt burden
  - Original: whatever the type of institution, the more a student borrows the greater the chance of default (Choy & Li, 2006; Dynarski, 1994; Lochner & Monge-Naranjo, 2004)
  - Credit recovery: the likelihood of rehabilitated loan default increases as the loan amount increases, while the likelihood of repay does not affected by the loan amount

- Delinquent period
  - Credit recovery: as delinquent period of the original loan increases, the likelihood of rehabilitated loan default decreases and the likelihood of successful repay increases

Results

☐ Comparing with Original Loan Defaults

- Type of the student loan
  - Credit recovery: income contingent loan borrowers are more likely to be failed in their rehabilitation programs, compared to ordinary student loan borrowers

- Legal action
  - Credit recovery: if there were any legal actions before rehabilitation, those borrowers are more likely to terminate their rehabilitation program earlier (through default or repay) than others

- Debt relief
  - Credit recovery: if there were any debt relief in the rehabilitation program, those borrowers are more likely to successfully repay and less likely to be defaulted again

Results

☐ Comparing with Original Loan Defaults

- Type of postsecondary institution
  - Credit recovery:
    - borrowers in two-year postsecondary institutions are more likely to default their rehabilitated loans compared to borrowers in four-year postsecondary institutions
    - borrowers in graduate schools are more likely to terminate their rehabilitation program earlier (through default or repay) than borrowers in four-year postsecondary institutions

- Program of study
  - Original: the relationship between program of study and original loan default is less clear in the literature
  - Credit recovery: borrowers majored in engineering, science, medicine and pharmacy are less likely to default and more likely to repay their rehabilitated loans compared to borrowers majored in other areas

Conclusion

☐ Main Findings and Implications

- This study forecasts failure of student loan credit recovery program participants using machine learning algorithms with account-level dataset
  - Loan defaulters are not for conventional credit assessments
  - However, credit recovery programs need to predict participant's ability to repay
- We find that the artificial neural networks algorithm performs best in predicting credit risks of student loan defaulters
  - Logistic regression can be an efficient alternative considering the computational resources
- However, the information currently used in student loan credit recovery programs may not be sufficient for assessing bad debtors' ability to recover their credit

☐ Future Works

- Hyperparameters fine-tuning
- Dimension reduction

References

- **References**

  - Gross, J. P., Cekic, O., Hossler, D., & Hillman, N. (2009). What Matters in Student Loan Default: A Review of the Research Literature. Journal of Student Financial Aid, 39(1), 19-29.
  - Guo, R., Zhao, Z., Wang, T., Liu, G., Zhao, J., & Gao, D. (2020). Degradation state recognition of piston pump based on ICEEMDAN and XGBoost. Applied Sciences, 10, 6593. 10.3390/app10186593.
  - Han, B. S., Kang, H. G., & Jun, S. G. (2015). Student loan and credit risk in Korea. Economics Letters, 135, 121-125.
  - Looney, A., & Yannelis, C. (2019). How useful are default rates? Borrowers with large balances and student loan repayment. Economics of Education Review, 71, 135-145.
  - Mueller, H. M., & Yannelis, C. (2019). The rise in student loan defaults. Journal of Financial Economics, 131(1), 1-19.