Predictive Models for Credit Card Default

Peijie Xie Meixin Liu Ben Liu

University of Victoria

April 3, 2018

Table of Contents

Table of Contents

Background

It is a necessity for banks to construct models to decide whether a client would default his credit card payment.

- Credit risk is the traditional risk of banking industry and managing credit risk has been a key part of the banking business.
- To increase market share, the credit card issuers over-issued credit cards to many unqualified applicants.
- Card holders overused credit card regardless of their repayment ability.
- It would be too expensive have a close background check for each individual applicants

Table of Contents

Data

Payment data from a bank in Taiwan in October, 2005

Variable Name	Description
Y	Default status
$\mathrm{LIMIT}_{\mathrm{BAL}}$	Credit limit
SEX	Gender
EDUCATION	Education
MARRIAGE	Marital status
AGE	Age
$\mathrm{PAY}_0 - \mathrm{PAY}_6$	Historical payment
$\mathrm{BILL}_{\mathrm{AMT1}} - \mathrm{BILL}_{\mathrm{AMT6}}$	Bill statement amount
$\mathrm{PAY}_{\mathrm{AMT1}} - \mathrm{PAY}_{\mathrm{AMT6}}$	Previous payment amount

Outlier Detection

- Use Mahalanobis distance to detect outliers
- 95% quantile from the chi-square distribution

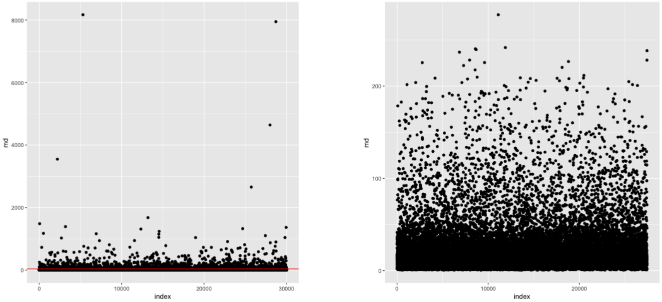


Figure: Mahalanobis Distance

Dimension Reduction

- There is a strong linear relationship between some predictors
- Collinearity will cause problems for our later model fitting

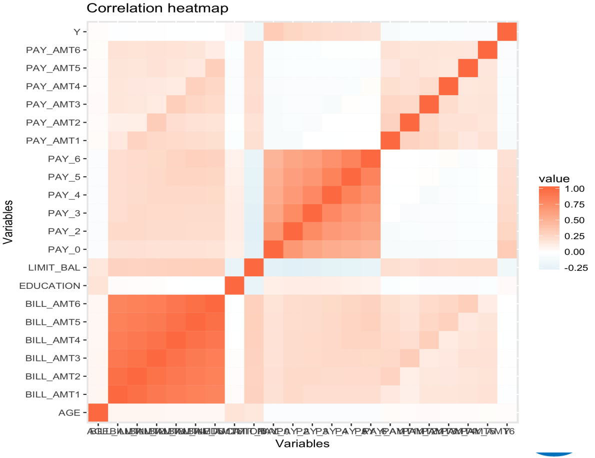
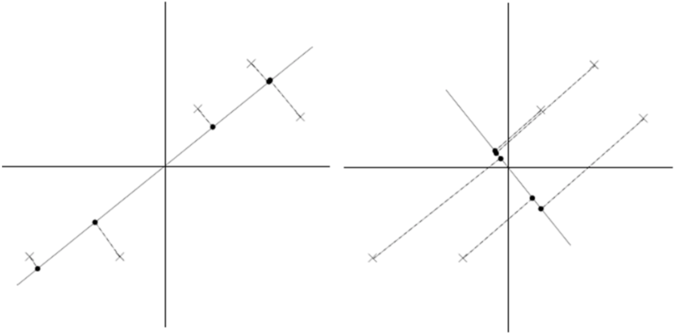


Figure: Correlation Heatmap for Predictors

Principal Component Analysis

- Principal component: Linear combination of inputs
- Aim: Capture as much of the variation as possible



Principal Component Analysis

Plot of variance (Scree plot) for each PC

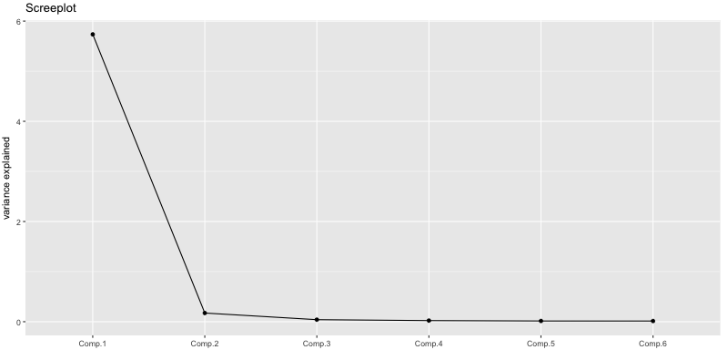


Table of Contents

Evaluation Criteria

We use the following criteria to compare our models.

- Accuracy
- ROC Curve
- Lift Chart

Confusion Matrix

		Actual Class	
		Positive(1)	Negative (0)
Predicted Class	Positive(1)	TP	FP
	Negative(0)	FP	TN
		Sensitivity	1-Specificity
		$TPR = \frac{TP}{TP+FN}$	$FPR = \frac{FP}{FP+TN}$

Table: Table of Confusion

Accuracy $= \frac{TP+TN}{P+N}$,
where $P = TP + FN$ and $N = FP + TN$

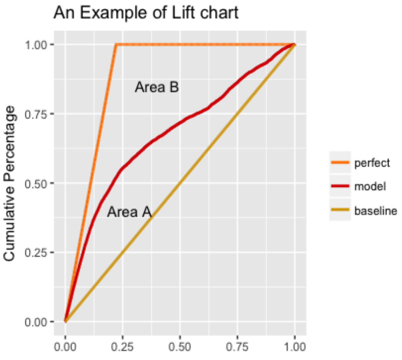
Brief Introduction to ROC

Given a set of prediction and a ranking classifier:

- Order the prediction by the probability from the highest to the lowest
- start from (1-FPR=0,TPR=0)
- for each prediction x in the sorted order
  if x is True Positive, move one step up
  if x is True Negative, move one step right
- evaluate AUC, where

$$AUC_{ROC} = \int_0^1 \frac{TP}{P} d\frac{FP}{N}$$

Lift Chart



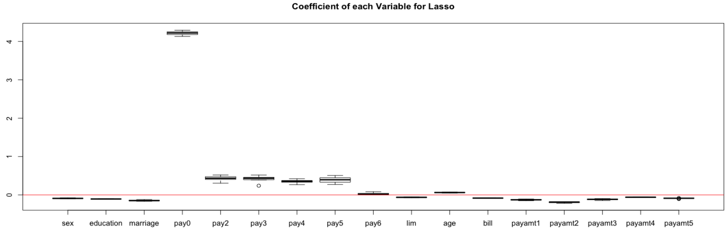
- A twist of ROC curve
- Depends on P:N ratio
- y-axis: TPR
- x-axis: $\frac{TP+FP}{P+N}$
- Baseline represents random predictions.
- Prefect line represents predictions without misclassification.
- Comparing criterion: Accuracy-ratio, where Accuracy-ratio$=\frac{Area_A}{Area_A+Area_B}$

Table of Contents

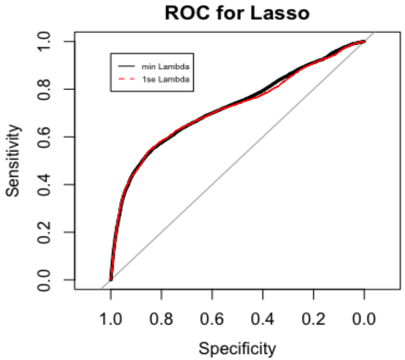
Fit Lasso, Ridge and Elastic net model

- 10-fold Cross-Validation
- For Lasso, Ridge and Elastic net, fit each model with 1se and min lambda.
- When fitting elastic net, use second layer of CV to decide weight between L1 and L2 penalty.

Compare the Importance of each Variable by its Coefficient

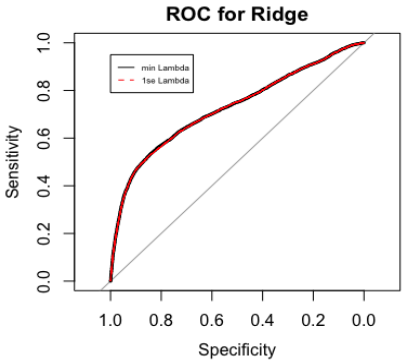


ROC curve for Lasso and AUC



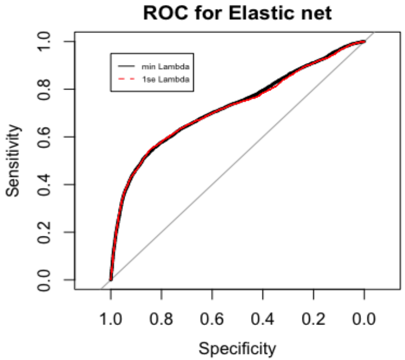
- Sensitivity = TPR
- Specificity = 1-FPR
- AUC (Lasso with min lambda) = 0.7252
- AUC (Lasso with 1se lambda) = 0.7217

ROC curve for Ridge and AUC



- Sensitivity = TPR
- Specificity = 1-FPR
- AUC (Ridge with min lambda) = 0.726
- AUC (Ridge with 1se lambda) = 0.7256

ROC curve for ELastic net and AUC



- Sensitivity = TPR
- Specificity = 1-FPR
- AUC (Elastic net with min lambda) = 0.7254
- AUC (Elastic net with 1se lambda) = 0.723

Compare Area Under Curve (AUC)

Table: AUC table

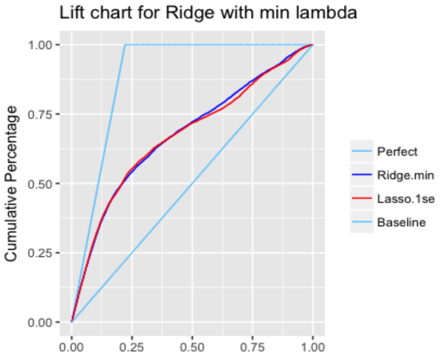
Model	AUC
Ridge with min Lambda	0.726
Ridge with 1se Lambda	0.7256
Lasso with min Lambda	0.7252
Lasso with 1se Lambda	0.7217
Elastic net with min Lambda	0.7254
ELastic net with 1se Lambda	0.723

Compare Accuracy Rate (With the best threshold)

Table: Accuracy Rate table

Model	Threshold	Accuracy
Ridge with min Lambda	0.4134485	0.8170759
Ridge with 1se Lambda	0.4101963	0.8160653
Lasso with min Lambda	0.4229622	0.8178245
Lasso with 1se Lambda	0.4069446	0.8192843
Elastic net with min Lambda	0.4257687	0.8180491
ELastic net with 1se Lambda	0.4063384	0.8173379

Lift Chart



- Accuracy-ratio (Ridge with min lambda) $= 0.4519638$
- Accuracy-ratio (Lasso with 1se lambda) $= 0.4434366$

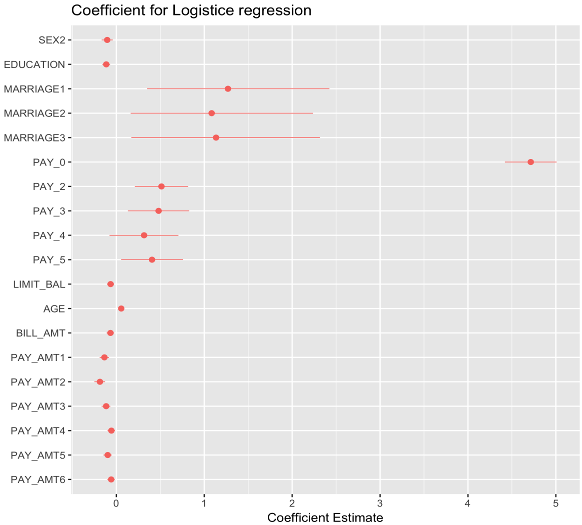
Logistic Regression

Basic Idea:

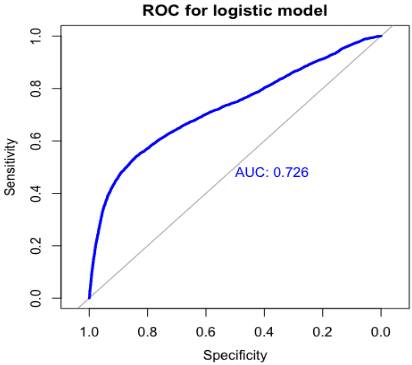
- Logistic regression is the type of regression we use for a response variable (Y) that follows a binomial distribution
- Model the log odds of the event (in our case, default) as a function of predictor variables

Logistic Regression

- Coefficient for the Logistic regression after variable selection (AIC)



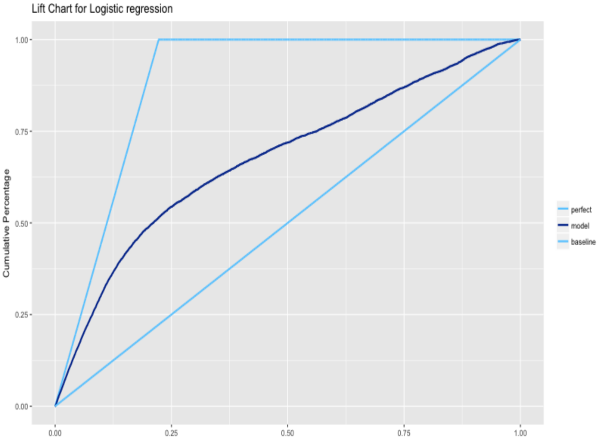
ROC Curve for Logistic Regression



- Sensitivity = TPR
- Specificity = 1-FPR
- AUC = 0.726
- optimal cutoff = 0.4369
- accuracy rate = 0.817

Logistic Regression

- Lift Chart for Logistic Regression
- Accuracy Ratio $= 0.4521$



Machine Learning Models

Non-parametric models:

- K-Nearest Neighbors
- Tree methods: Decision Tree and Random Forest

Parametric models:

- Linear Discriminant Analysis
- Naive Bayes

Machine Learning Models

Non-parametric models:

- K-Nearest Neighbors
- Tree methods: Decision Tree and Random Forest

Parametric models:

- Linear Discriminant Analysis
- Naive Bayes

Criteria:
Area Under Curve (AUC) and Accuracy

Table of Contents

K-Nearest Neighbors

- Select a positive integer k
- Select closest k points near the new sample
- Find most common classification

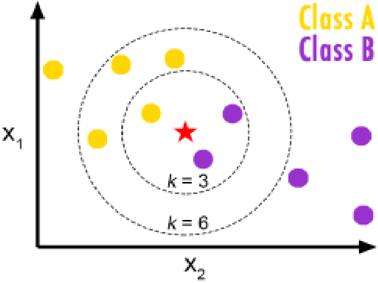


Figure: KNN Algorithm

K-Nearest Neighbors

- Select a positive integer k
- Select closest k points near the new sample
- Find most common classification

How do we select k?

- Try $k = 1, 2, \cdots, 100$
- 10-fold cross validation

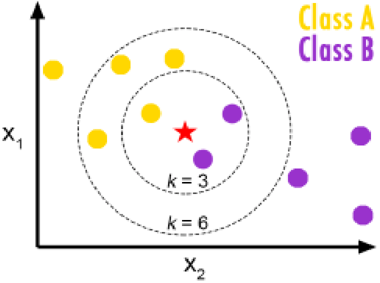


Figure: KNN Algorithm

K-Nearest Neighbors

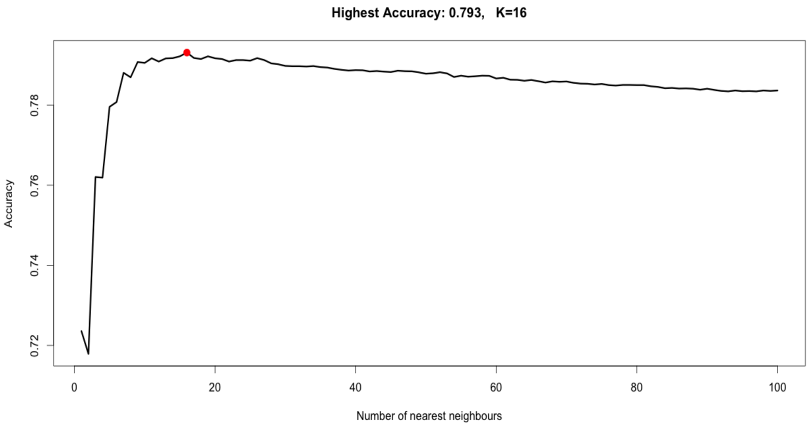


Figure: Tuning parameter: k

K-Nearest Neighbors

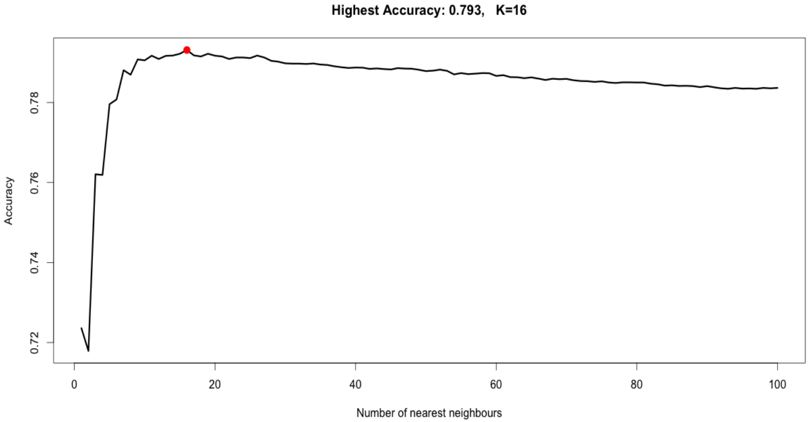


Figure: Tuning parameter: k

AUC : 0.547

Tree Methods

Decision Tree:

- Visual representation
- Accommodate non-linear relations
- Easy to be interpreted!

Decision Tree

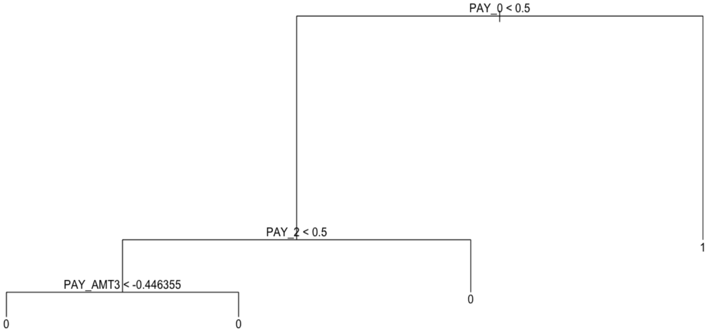


Figure: Decision Tree Before Pruning

Decision Tree

```
          CP nsplit rel error    xerror       xstd
1 0.1853609      0 1.0000000 1.0000000 0.01148927
2 0.0100000      1 0.8146391 0.8146391 0.01063895

Variable importance
PAY_0 PAY_5 PAY_2 PAY_6 PAY_4 PAY_3
   93     2     1     1     1     1
```

Figure: Complexity Parameter

Decision Tree

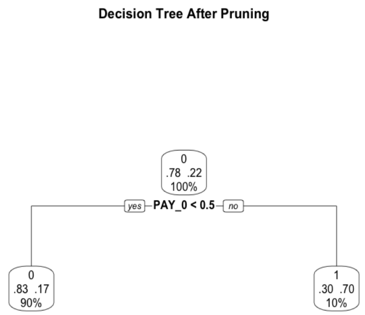
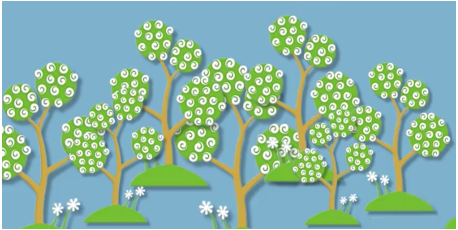


Figure: Decision Tree After Pruning

Tree Methods

Random Forest:

- More robust than decision tree using bootstrap
- Randomly select observations and predictors
- Majority rule



Random Forest

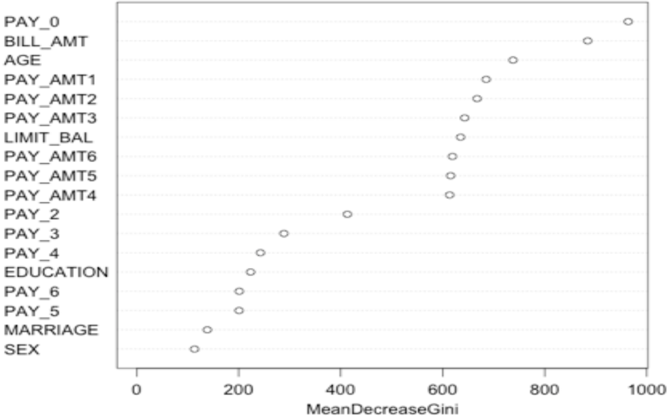


Figure: Importance of Variables

Random Forest

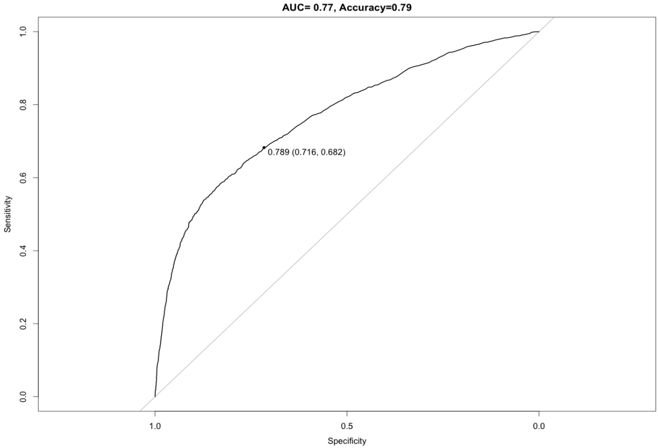


Figure: Random Forest: AUC

Random Forest

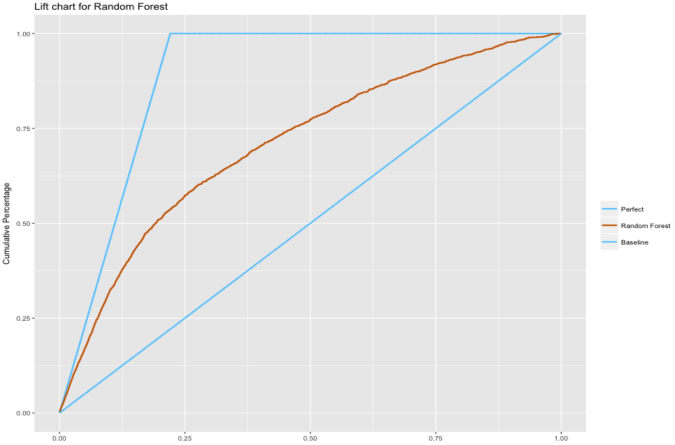


Figure: Accuracy ratio: 0.549

Parametric models

Linear Discriminant Analysis:

- Commonly used method to reduce high dimensions
- Find boundaries around clusters of classes

Naive Bayes:

- Probabilistic classifiers based on applying Bayes' theorem

Parametric models

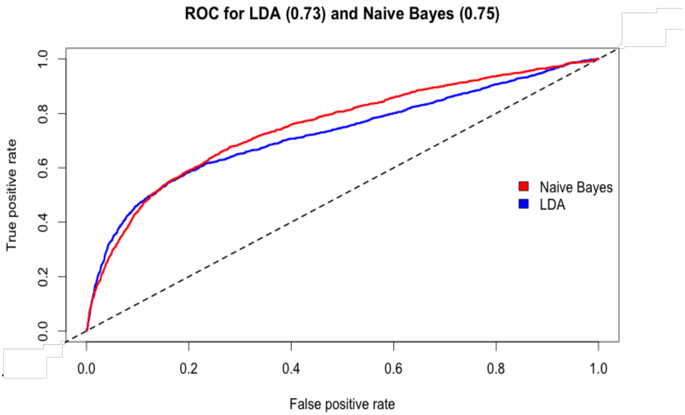


Figure: LDA v.s. Naive Bayes

Model Comparison

Table: Best Model Comparison

Model	AUC	Accuracy
Random Forest	0.77	0.787
KNN	0.54	0.793
Logistic	0.73	0.816
Ridge (min)	0.73	0.817
Lasso (1se)	0.72	0.819
LDA	0.73	0.812
Naive Bayes	0.75	0.810

Table of Contents

Conclusion

Table: Model Comparison

Model	Accuracy Ratio
Random Forest	0.539
Ridge (min)	0.452
Lasso (1se)	0.443
Logistic	0.452

Conclusion

Table: Model Comparison

Model	Accuracy Ratio
Random Forest	0.539
Ridge (min)	0.452
Lasso (1se)	0.443
Logistic	0.452

Random forest wins!

Reference

[1] Baesens B, Van Gestel T, Viaene S, Stepanova M, Suykens J, Vanthienen J. Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the operational research society. 2003 Jun 1;54(6):627-35.

[2] Härdle W, Simar L. Applied multivariate statistical analysis. Berlin: Springer; 2007 Aug 9.

[3] James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. New York: springer; 2013 Feb 11.

[4] Yeh IC, Lien CH. The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications. 2009 Mar 1;36(2):2473-80.

[5] Data source https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset

Thank You