

DEFAULT OF CREDIT CARD MODELING AND VISUALIZATION

GROUP 8 :

JINGTIAN YAO

CHENFEI JIANG

MICHAEL UTOMO

CATEGORY

1. Problem Description

2 Data Preprocessing

3. Model Selection

4. Visualization

PROBLEM DESCRIPTION

The accuracy of the result of customers' credit card default payment is essential from the perspective of **risk management**.

We aimed at predicting customers' default payments status next month and compares the predictive accuracy among different algorithms.

We also built a **Shiny App** for the bank to help them with better risk management.

DATA PREPROCESSING

- Data Source :
<https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients>
- **23 attributes**
- Amount of given credit
- Gender
- Education
- Marital Status
- Age
- History of past payment (last 6 months)
- Amount of bill statement (last 6 months)
- Amount of previous payment (last 6 months)

Divide data into training set (**24,000 observations**) and test set (**6,000 observations**)

MODEL SELECTION

1. SVM (radial basis function kernel)

- Tuning parameters: **cost** & **gamma**

range of cost: $10^{(-6:2)}$

range of gamma: $10^{(-4:4)}$

- Best parameters: **cost = 10, gamma = 0.001**

For each algorithm, use 5-folds cross validation to choose the best parameter.

MODEL SELECTION

2. Random Forest

- Tuning parameters : **mtry** & **ntree**

range of mtry : **1,2,3,4**

range of ntree : **100, 200, 300, ..., 1000**

- Best parameters: **mtry = 2, ntree = 600**

For each algorithm, use 5-folds cross validation to choose the best parameter.

MODEL SELECTION

3. Xgboost

- Tuning parameters : eta & gamma & max depth

range of eta : 0.3, 0.5, 0.7

range of gamma : 0, 0.5, 1

range of max depth : 2, 3, 4, ..., 10

Best parameters: eta = 0.5, gamma = 0.5, max depth = 4

For each algorithm, use 5-folds cross validation to choose the best parameter.

MODEL SELECTION

4. Logistic Regression

	Variables	Formula	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	AUC	P < 0.05
2474	2	Y ~ X21 + X16	X21	X16	0	0	0	0	0	0	0	0	0.5731454	0
2056	2	Y ~ X19 + X14	X19	X14	0	0	0	0	0	0	0	0	0.5820459	0
2483	2	Y ~ X21 + X17	X21	X17	0	0	0	0	0	0	0	0	0.5825252	0
2274	2	Y ~ X20 + X15	X20	X15	0	0	0	0	0	0	0	0	0.5834993	0
2367	2	Y ~ X21 + X4	X21	X4	0	0	0	0	0	0	0	0	0.5836388	0
2547	3	Y ~ X21 + X22 + X13	X21	X22	X13	0	0	0	0	0	0	0	0.5837639	0
2665	3	Y ~ X22 + X21 + X13	X22	X21	X13	0	0	0	0	0	0	0	0.5837639	0
2067	2	Y ~ X19 + X15	X19	X15	0	0	0	0	0	0	0	0	0.5841984	0
2548	3	Y ~ X21 + X22 + X15	X21	X22	X15	0	0	0	0	0	0	0	0.5846775	0
2666	3	Y ~ X22 + X21 + X15	X22	X21	X15	0	0	0	0	0	0	0	0.5846775	0
2560	3	Y ~ X21 + X23 + X13	X21	X23	X13	0	0	0	0	0	0	0	0.5849772	0

AUC: **Area under the ROC curve** - assessing discrimination in logistic regression

For each algorithm, use 5-folds cross validation to choose the best parameter.

MODEL SELECTION

5. Neural Network

Size : **Number of units in the hidden layer** (Choose 2 from 1 to 5)

Decay : **parameter for weight decay** (0.1)

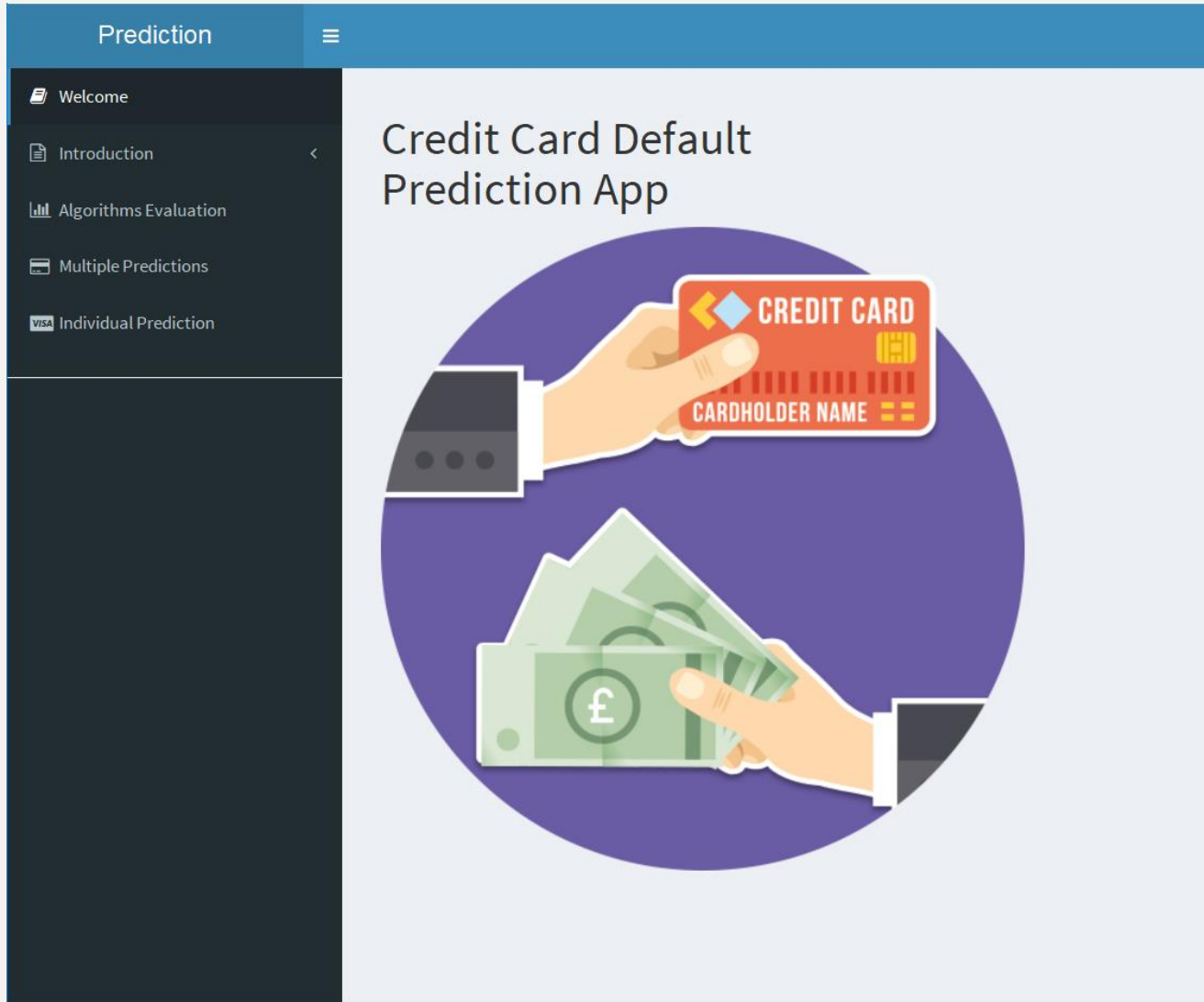
Maxit: **maximum number of iterations** (200)

For each algorithm, use 5-folds cross validation to choose the best parameter.

Algorithm	Test Accuracy	Prediction Time (6000 observations)
SVM	0.779	28.75s
Random Forest	0.815	3.07s
Xgboost	0.878	0.05s
Logistic Regression	0.808	0.10s
Neural Network	0.779	4.41s

MODEL SELECTION

According to the test accuracy, **Xgboost** algorithm performs the best.



VISUALIZATION

The homepage of our app.

The app includes 5 parts:

Welcome

Introduction

Algorithm Evaluation

Multiple Predictions

Individual Prediction

Welcome

Introduction

Algorithms Evaluation

Multiple Predictions

Individual Prediction

Choose Your Testing Data Table (CSV file)

Browse...

testforshow2.csv

Upload complete

Submit

Tutorial

Test Data

Prediction

To make multiple prediction, a data frame consisting the relevant information of credit card clients you want to predict is required. The data frame should include 16 explanatory variables as follows.

X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.

X2: Gender (1 = male; 2 = female).

X3: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).

X4: Marital status (1 = married; 2 = single).

X5: Age (year).

X6-X11: History of past 6 months payment. (-1 = pay duly; 1~8 = payment delay for 1~8 months; 9 = payment delay for 9 months or above).

X12-X17: Amount of bill statement (NT dollar) of the past six months.

X18-X23: Amount of payment (NT dollar) of the past six months.

Example:

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23
1	230000	1	1	2	35	2	2	2	2	2	2	117277	119823	120988	121464	123740	120937	5800	4500	3800	5000	0	4500
2	130000	1	3	1	56	-1	-1	-1	0	-1	-1	582	0	582	291	441	441	0	582	0	441	441	291
3	50000	2	2	1	31	2	0	0	0	2	0	0	43979	44970	48101	46993	47940	50903	1724	3855	0	1694	3713
4	30000	2	2	2	25	0	0	0	0	0	0	-2	15493	16427	17379	23100	0	0	1500	1279	10000	0	0
5	220000	2	1	2	35	-1	-1	-1	-1	-1	-1	5564	7443	5572	5572	5572	5774	7479	5600	5600	5600	5802	3279

VISUALIZATION

Welcome

Introduction

Algorithms Evaluation

Multiple Predictions

Individual Prediction

Choose Your Testing Data Table (CSV file)

Browse...

testforshow2.csv

Upload complete

Submit

Tutorial

Test Data

Prediction

Show 5 entries

Search:

X	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	PAY_5	PAY_6	BILL_AMT1
1	230000	1	1	2	35	2	2	2	2	2	2	117277
2	130000	1	3	1	56	-1	-1	-1	0	-1	-1	582
3	50000	2	2	1	31	2	0	0	2	0	0	43979
4	30000	2	2	2	25	0	0	0	0	0	-2	15493
5	220000	2	1	2	35	-1	-1	-1	-1	-1	-1	5564

X

LIMIT_BAL

SEX

EDUCATION

MARRIAGE

AGE

PAY_0

PAY_2

PAY_3

PAY_4

PAY_5

PAY_6

BILL_AMT1

Showing 1 to 5 of 6,000 entries

Previous

1

2

3

4

5

...

1200

Next

VISUALIZATION

Prediction

Welcome

Introduction

Algorithms Evaluation

Multiple Predictions

Individual Prediction

Choose Your Testing Data Table (CSV file)

Browse...

testforshow2.csv

Upload complete

Submit

TutorialTest DataPrediction

Show 25 entries

Search:

ID	Xgboost	Random_Forest
1	Yes	No
2	Yes	Yes
3	Yes	No
4	No	No
5	No	No
6	No	No
7	No	No
8	No	No
9	Yes	Yes

VISUALIZATION

Prediction

Welcome

Introduction

Algorithms Evaluation

Multiple Predictions

Individual Prediction

Gender

☐ Male ☒ Female

Marriage

☐ Married ☒ Unmarried

Education

☒ Graduate or above ☐ University ☐ High School ☐ Others

Amount of given credit

0

Age

30

History of past 6 months payment

Pay Duly

Pay Duly

Pay Duly

Pay Duly

Pay Duly

Pay Duly

Amount of bill statement last 6 months

0

0

0

0

0

0

Amount of payment last 6 months

0

0

0

0

0

0

Predict

Tutorial

Prediction

Will this customer make default payment next month?

No

VISUALIZATION

https://lanceyjtcolumbia.shinyapps.io/default_payment/

FUTURE PLAN

- Add “**Model Training**” feature in our app.

Idea: Allow user to **upload their own training dataset**, and we can train a **new model** for the user based on the data user given.



Q & A



THANK YOU