Machine Learning

Telecom Customer Churn Prediction Analysis

Group 14:

Xuewei Zhang
Xiaokai Ma
Jue Gu

Business Overview

Objective:

Predict the Churn

Consumer Behavior : I

"Good "Phase

Consumer Behavior : II

"Action "Phase

Consumer Behavior : III

"Churn "Phase

Data Preparation

01

Data Source:
Kaggle: # telecom churn

```
[ ]  df.shape

     (99999, 226)
```

	churn
0	92.17%
1	7.83%

02

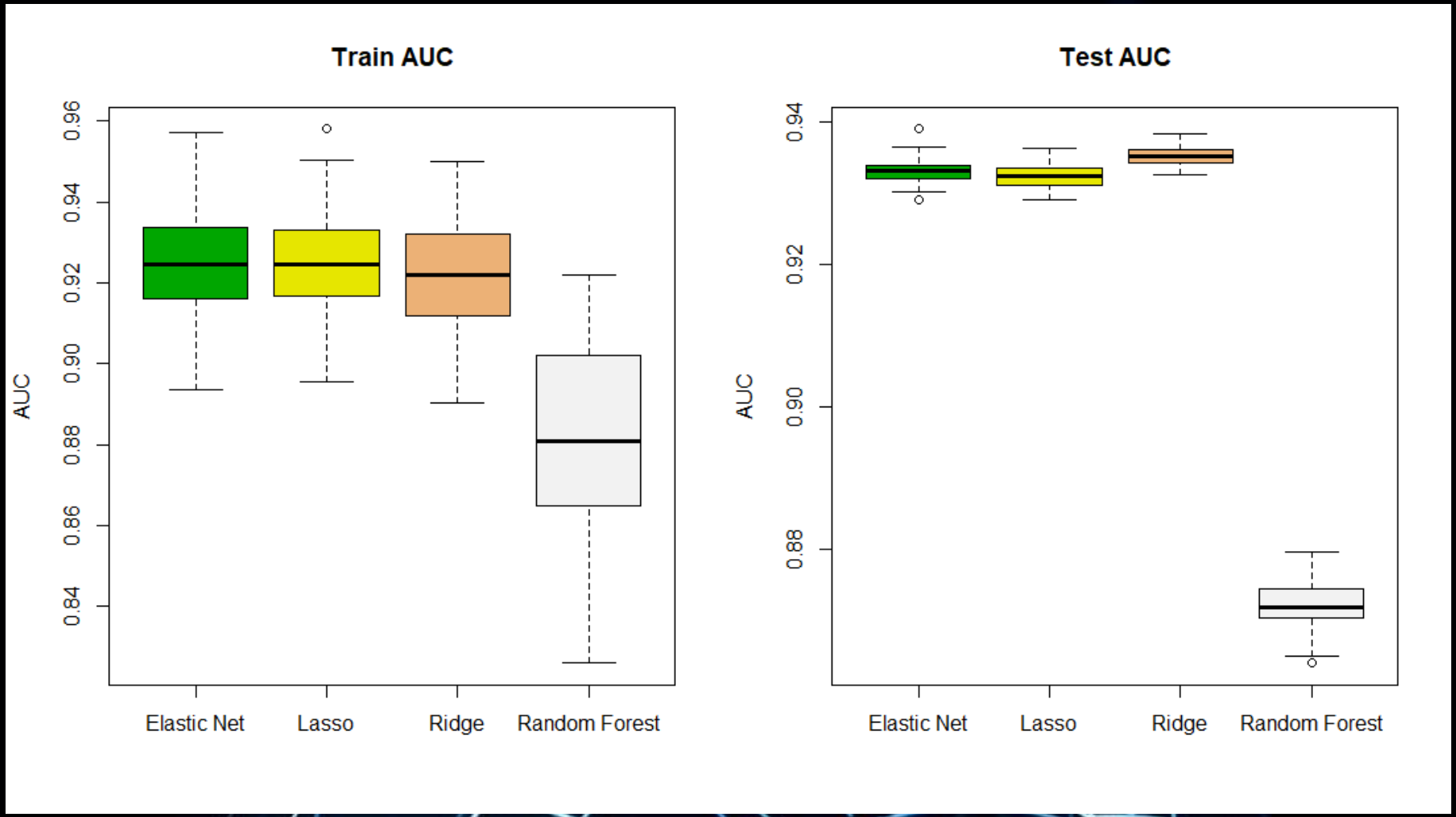
Filter high-value customers
Impute missing value
Identify redundant features

Identify Churn Customers
Imbalance Data

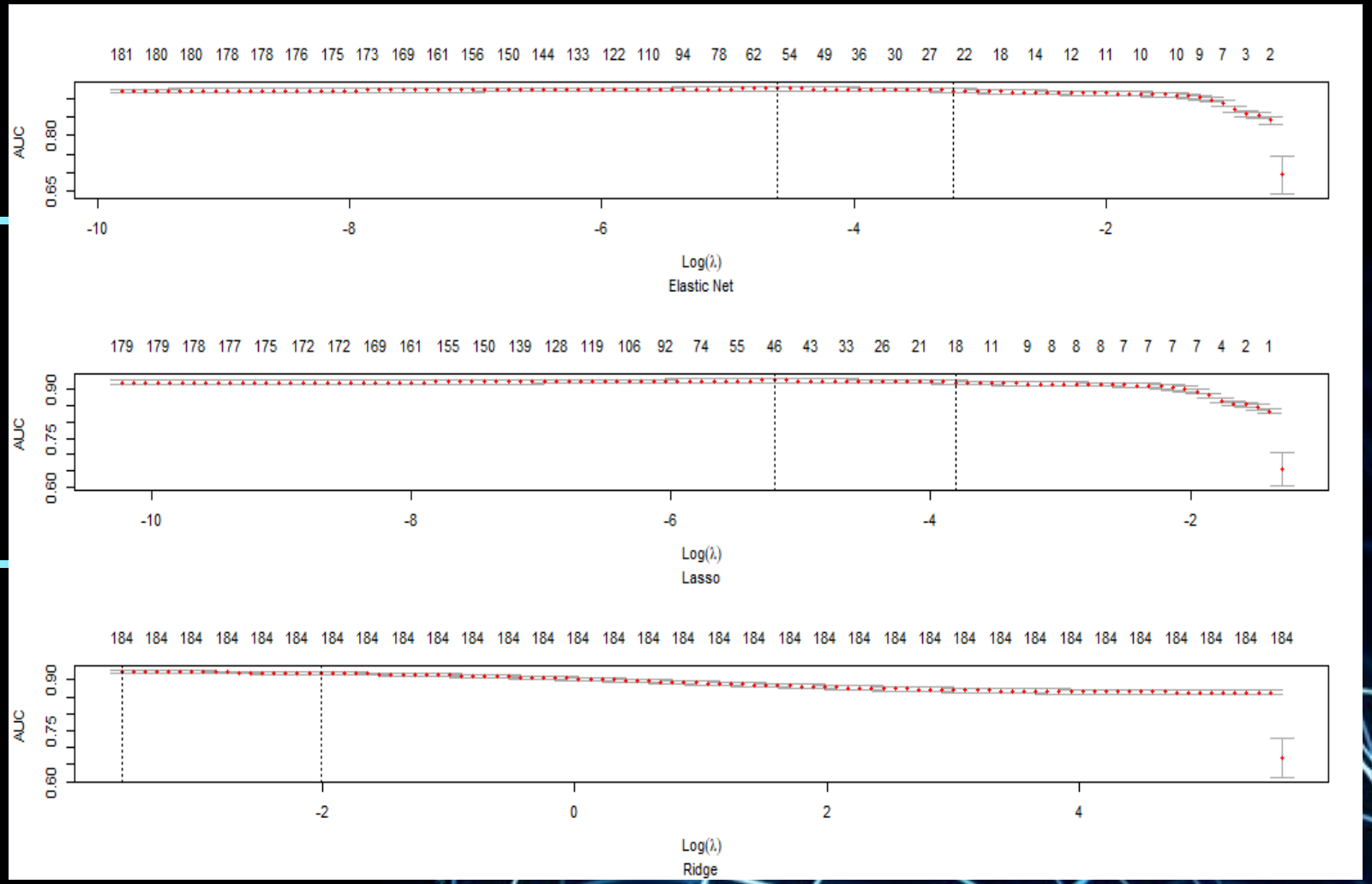
```
[ ]  df.shape

     (15014, 184)
```

A U C


Train AUC
AUC
0.84
0.86
0.88
0.90
0.92
0.94
0.96
Elastic Net
Lasso
Ridge
Random Forest
Test AUC
AUC
0.88
0.90
0.92
0.94
Elastic Net
Lasso
Ridge
Random Forest

10-fold CV curves



Elastic: 1.72mins

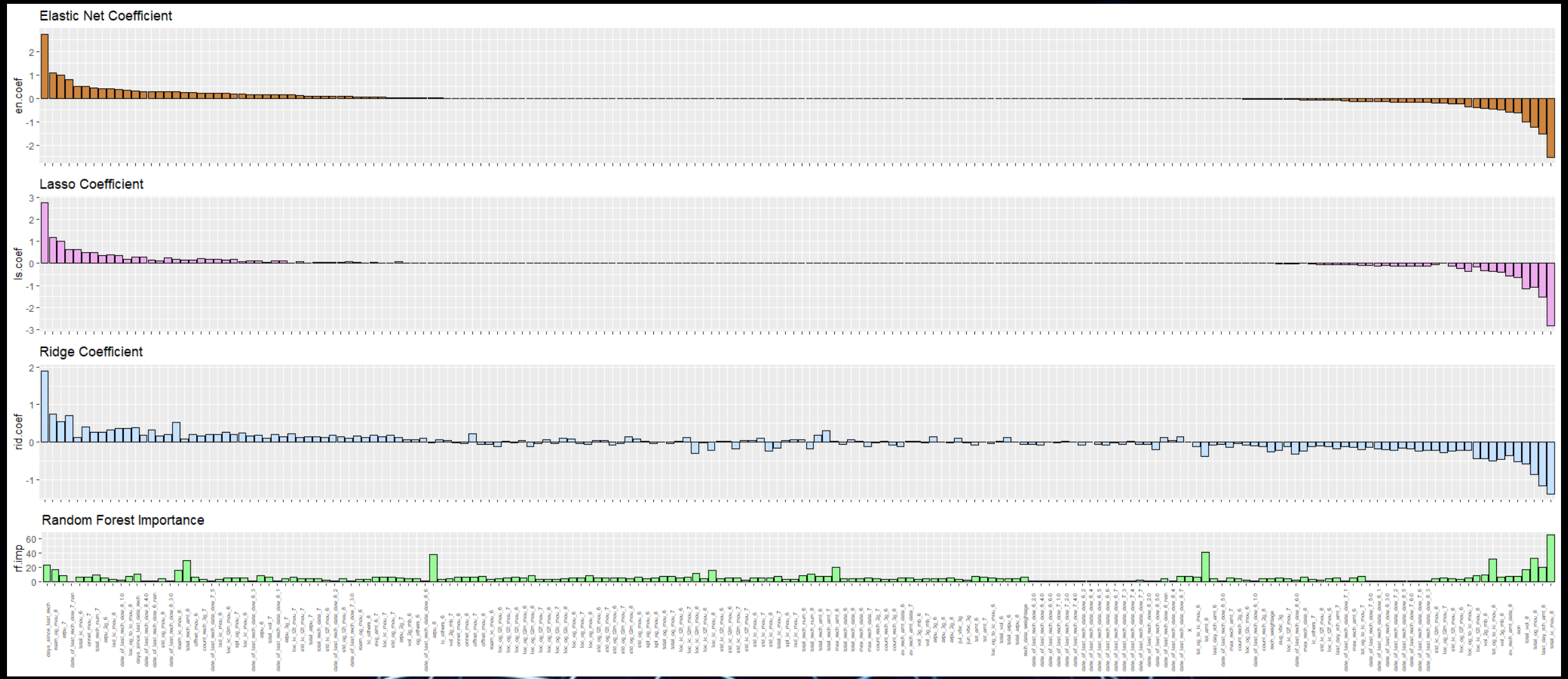
Lasso: 1.71mins

Ridge: 0.79 mins

Trade off

	Test AUC	Time for single
Elastic	0.934	1.808(min)
★ Ridge	0.935	0.8160(min)
Lasso	0.933	1.8135(min)
Random Forest	0.881	3.78(min)

Coefficients



Feature Importance Rank

Rank	Features	Elastic Net
1	days_since_last_rech	2.73
2	roam_og_mou_8	1.08
3	arpu_7	0.98
4	date_of_last_rech_dow_7_nan	0.79
5	total_ic_mou_6	0.52
6	onnet_mou_7	0.51
7	total_rech_num_7	0.44
8	arpu_2g_6	0.41
9	isd_ic_mou_8	0.40
10	date_of_last_rech_dow_8_1.0	0.39
11	loc_og_to_ic_mou_8	0.34
12	days_since_last_data_rech	0.32
13	date_of_last_rech_dow_8_4.0	0.29
14	date_of_last_rech_dow_6_nan	0.28
15	std_og_mou_8	0.28
16	date_of_last_rech_dow_8_3.0	0.27
17	roam_ic_mou_8	0.27
18	total_rech_amt_8	0.24
19	offnet_mou_6	0.24
20	count_rech_3g_7	0.22

Rank	feature	Lasso
1	days_since_last_rech	2.75
2	roam_og_mou_8	1.19
3	arpu_7	1.00
4	date_of_last_rech_dow_7_nan	0.64
5	total_ic_mou_6	0.63
6	total_rech_num_7	0.49
7	onnet_mou_7	0.48
8	isd_ic_mou_8	0.39
9	arpu_2g_6	0.37
10	date_of_last_rech_dow_8_1.0	0.37
11	days_since_last_data_rech	0.30
12	date_of_last_rech_dow_8_4.0	0.27
13	date_of_last_rech_dow_8_3.0	0.24
14	count_rech_3g_7	0.22
15	isd_ic_mou_6	0.20
16	date_of_last_rech_data_dow_7_5	0.19
17	loc_og_to_ic_mou_8	0.18
18	roam_ic_mou_8	0.18
19	spl_og_mou_7	0.18
20	offnet_mou_6	0.16

Rank	Features	Ridge
1	days_since_last_rech	1.90
2	roam_og_mou_8	0.76
3	date_of_last_rech_dow_7_nan	0.71
4	arpu_7	0.55
5	roam_ic_mou_8	0.52
6	onnet_mou_7	0.40
7	days_since_last_data_rech	0.38
8	date_of_last_rech_dow_8_1.0	0.37
9	loc_og_to_ic_mou_8	0.36
10	isd_ic_mou_8	0.32
11	date_of_last_rech_dow_6_nan	0.32
12	total_rech_amt_7	0.30
13	total_rech_num_7	0.27
14	loc_ic_t2m_mou_6	0.27
15	arpu_2g_6	0.27
16	loc_ic_mou_6	0.24
17	loc_ic_t2t_mou_7	0.22
18	offnet_mou_7	0.22
19	date_of_last_rech_data_dow_7_5	0.21
20	spl_og_mou_7	0.20

Rank	Features	Random Forest Importance
1	total_ic_mou_8	65.27
2	tot_amt_8	41.31
3	arpu_8	37.95
4	total_og_mou_8	32.58
5	tot_og_to_ic_mou_8	32.09
6	total_rech_amt_8	28.98
7	days_since_last_rech	22.51
8	max_rech_amt_8	19.57
9	last_day_rch_amt_8	19.39
10	roam_og_mou_8	16.32
11	total_vol_8	16.12
12	roam_ic_mou_8	15.28
13	loc_ic_mou_8	15.02
14	loc_ic_t2m_mou_8	11.56
15	days_since_last_data_rech	10.60
16	total_rech_num_8	10.41
17	total_rech_num_7	8.86
18	vol_2g_mb_8	8.85
19	loc_og_t2m_mou_8	8.17
20	loc_ic_t2t_mou_8	8.02

Insights

k

Imbalance

Trade off

Multicollinearity

Feature Importance

THANK YOU