Optimizing Operation Room Utilization by Predicting Surgery Duration

Project Team 4

102034606 WU, CHOU-CHUN

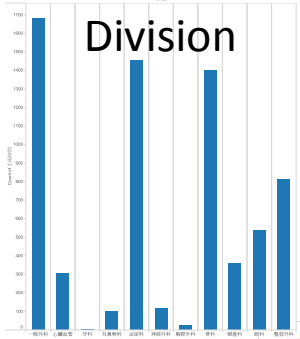
103078508 CHEN, LI-CHAN

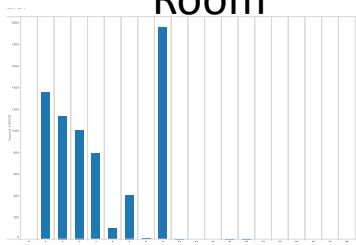
102077503 LI, DAI-SIN

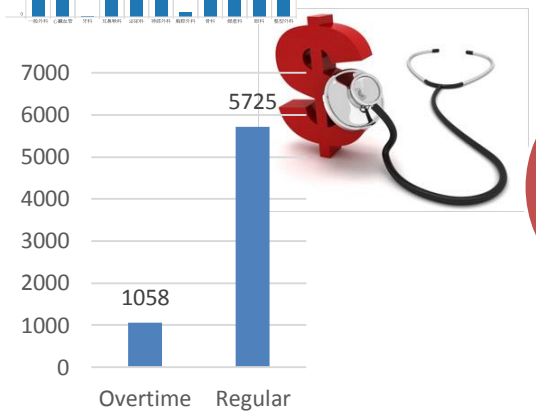
Advisor: Galit Shmueli

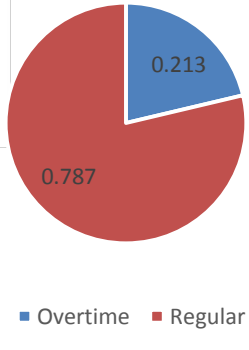
Business Goal

Conflict of Interest on Scheduling







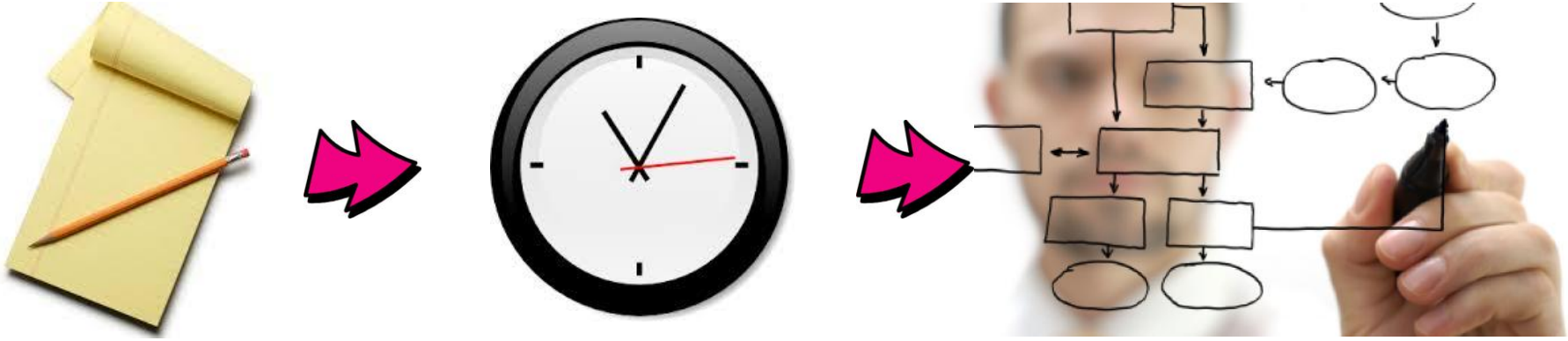


uneven distribution to the usage

establishment of an efficient surgery scheduling

Data Mining Goal

Supervised Data mining problem



Predict the total operation time

Better utilize in the scheduling

Data description

There are 2425 columns and 6783 rows

Numerical	
Operation Time	in minutes
Age	in years
Categorical (levels)	
Room#	20
Division	11
Gender	2
TreatType	315(67)
SurgeryType	2
AnesthesiaType	12
Doctor	43
Assistant1	26
Assistant2	10
Diagnosis1	744(67)
Diagnosis2	406(67)
Method1	470(67)
Method2	275(67)
ScrubbingNurse1	26
ScrubbingNurse2	26
CirculatingNurse1	20
CirculatingNurse2	15

- There are originally 19 columns on the raw data set.
- The data includes surgery records from a hospital in Taipei during the years 2010 and 2011.
- Partitioning: randomly choose 50% for training and 50% for testing
- As suggested, we bin the variables with many levels according to the dependent variable (can be reduced to 550)

Methods

- Explore and pre-process the data
  - Missing value
  - Data transformation
- Create Dummy/ Binning
- Variable Selection
- Model Construction
  - ✓Linear Regression(Lasso)
  - ✓Naïve Bayes
  - ✓Regression Tree
  - ✓Random Forest
- Cross Validation

$$\min_{\beta} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

Objective Function of
Simple Linear Regression

$$\min_{\beta} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

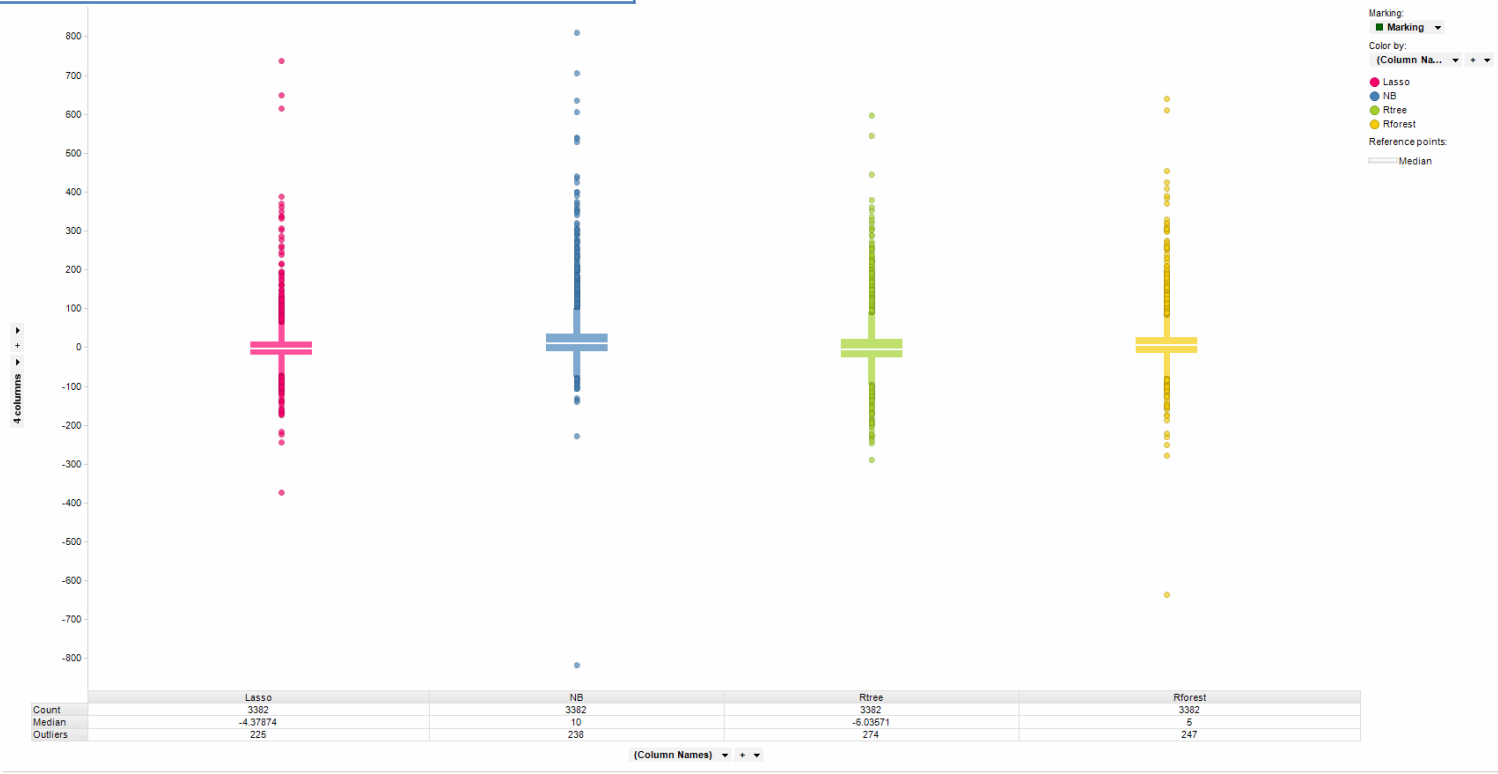
Objective Function of
Lasso Regression

Evaluation (1/2)

RMSE of Training Data			
Lasso	NB	RTree	RForest
30.723	63.502	44.322	2.047

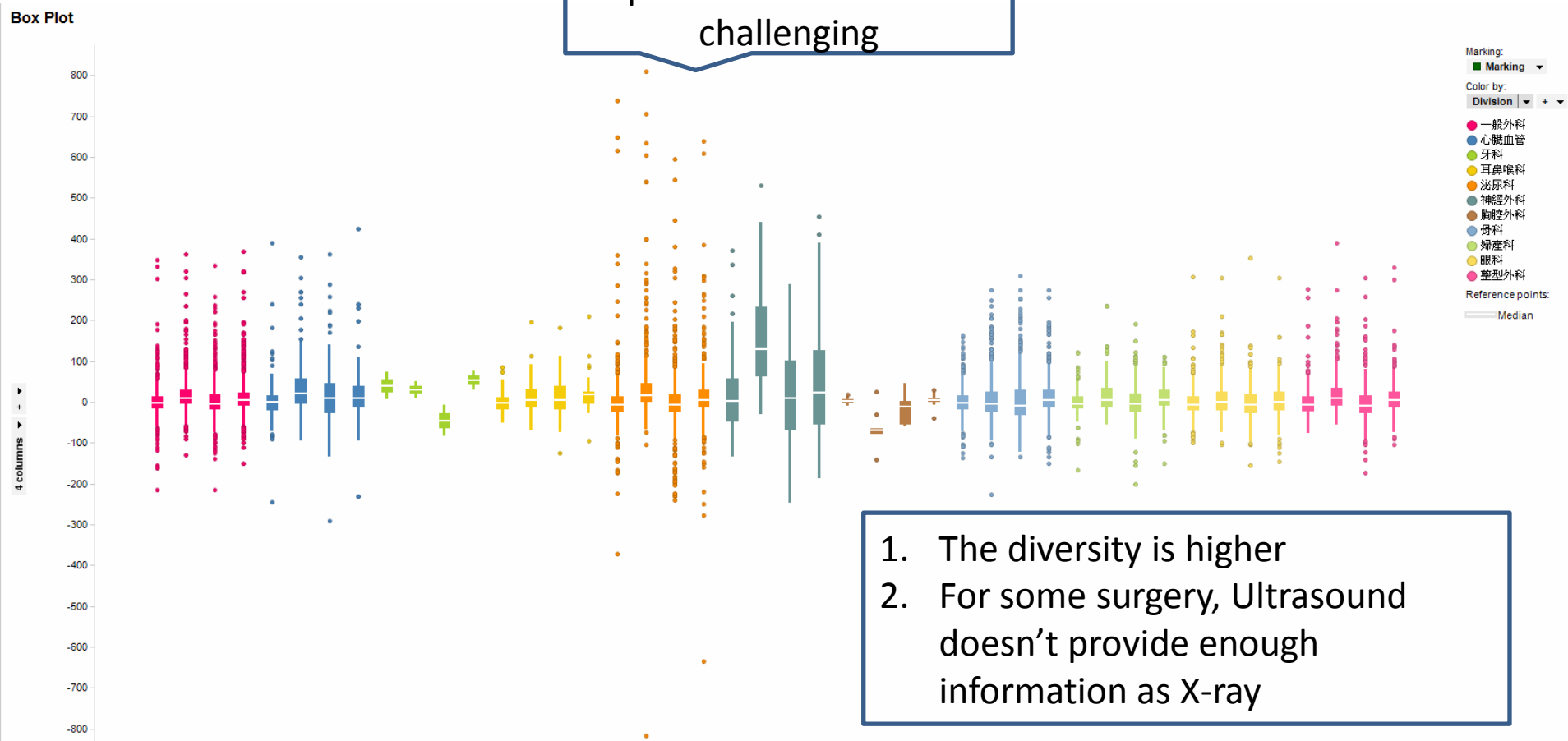
MAPE of Testing Data			
Lasso	NB	RTree	RForest
0.322	0.396	0.448	0.335

RMSE of Testing Data				
Lasso	NB	RTree	RForest	Naive
47.353	54.509	54.898	55.364	77.130



Evaluation (2/2)

The prediction of Urology operation time is more challenging



1. The diversity is higher
2. For some surgery, Ultrasound doesn't provide enough information as X-ray

	一般外科				心臓血管				牙科				耳鼻喉科				泌尿科				神経外科				胸腔外科				母科				婦産科				眼科				整型外科			
	Lasso	NB	Rtree	Rf...	La...	NB	Rtree	Rf...	La...	NB	Rtree	Rf...	La...	NB	Rtr...	Rf...	La...	NB	Rtr...	Rf...	La...	NB	Rtree	Rf...	La...	NB	Rtree	Rf...	La...	NB	Rtree	Rf...	La...	NB	Rtr...	Rf...	La...	NB	Rtr...	Rf...	La...	NB	Rtr...	Rforest
Count	827	827	827	827	134	134	134	134	2	2	2	2	46	46	46	46	764	764	764	764	60	60	60	60	11	11	11	11	686	686	686	686	180	180	180	180	272	272	272	272	400	400	400	400
Median	-3.11746	10	-5....	5	-0....	20	8.75	10	39....	30	-45...	52.5	-1....	5.5	4	18.5	-6....	15	-7	5	2.8...	130	8.1	22.5	1.1...	-70	-11	5	-3....	-5	-8....	5	-3....	5	-4....	5	-6....	0	-6.75	0	-7....	10	-8....	5
Outliers	52	57	72	64	14	9	9	6	0	0	0	0	2	2	2	5	43	52	73	57	4	1	0	2	1	3	0	2	43	41	27	36	16	7	13	10	18	18	11	12	13	18	26	22

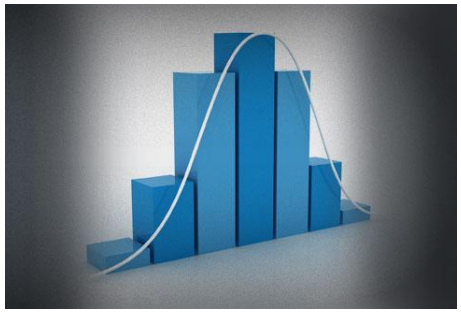
Division (Column Names)

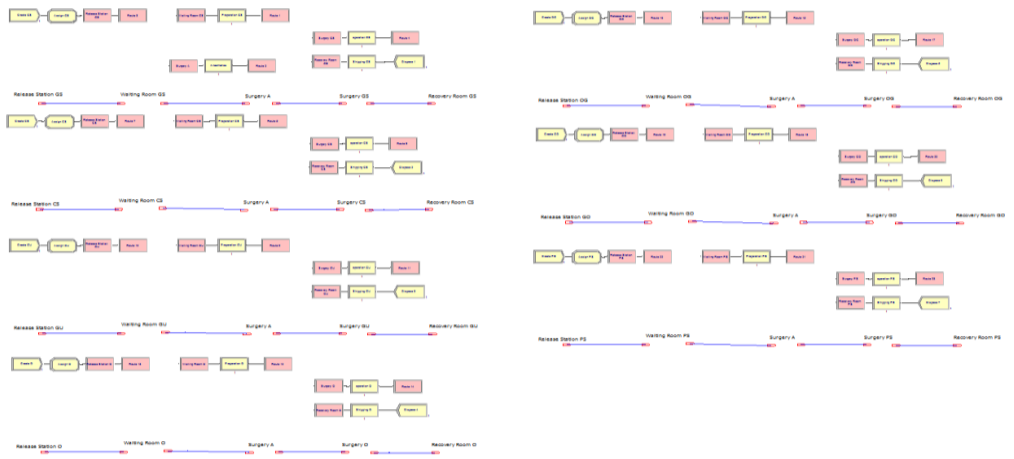
Recommendations (1/2)

- On average, the income of one surgery is around 8000NT and the overtime expense is 30 NT/min. Rule of thumb tells us that delaying the surgery for more than 3 hours will make a loss.
- It is costly to underestimate the operation time which may lead to possible overtime payment. Thus, from this perspective, the linear model should be an ideal choice.
- As for the practical usage on scheduling, we will generate some rules from the trees for further interpretation.

Recommendations (2/2)

- By controlling the waiting time (before entering the surgery room) within 15 minutes, the operation time can be reduced by 5%.
- According to the simulation result using Arena, by properly allocating the surgery room, the capacity can be increased by 7%.





Thank You!