Predicting surgical case duration for a Thorax Centre

Primary Topic: DPV, Secondary Topic: DM

Course: 2018-1A – Group: 80 – Submission Date: 2019-02-05

Feiyi Su  
 University of Twente  
 f.su@student.utwente.nl

Liyang He  
 University of Twente  
 l.he@student.utwente.nl

ABSTRACT

This paper provides a deep insight into surgical case durations for Thorax Centrum Twente (TCT). The exploration of project is based on the Data Science Course. The two main topics related to the project is data presentation and visualization (DPV) and data mining (DM).

The project aims to identify patterns in surgical case durations and to derive a predictive model in the end. In order to address the challenge, regression tree related knowledge is used to build predictive model and try to make the predictive model more accurate. Besides, relevant visualization results extracted from surgical data set are also provided.

KEYWORDS

Predictive Model, Data Mining, XGBoost, Gradient Boosting Decision Tree

1 INTRODUCTION

In modern healthcare, efficiency is directly linked with healthcare quality. Improving efficiency aims to provide more and better healthcare services. Especially when the modern society was scarce of medical resources, more and more hospitals are seeking for a way to improve efficiency so that they can ensure delivering better quality service with lower resources.

2 BACKGROUND

Our target organization is one thorax centre in MST. MST is a top clinical medical centre in Twente and is one of the biggest non-academic hospitals in the Netherlands. Thorax Centre Twente (TCT) is a centre within MST, which focuses on diagnosis and treatment of cardiothoracic diseases. TCT performs around more than 1,000 open-heart surgeries per year. Patient satisfaction is also a central theme for the TCT. Actually, the TCT gained high scores in patient satisfaction, as measured in national CQ standards [1]. TCT has been a huge progress since its establishment in September 2004. However, it still faces the problem of a high rate of operating rooms working beyond regular time. This problem causes low staff satisfaction and unnecessary resources waste. Therefore, the goal is to suggest more accurate predictions of surgical case duration to help OR-planner making most efficient OR-schedules to improve efficiency. Besides, the visualization of features helps to facilitate new discovery.

3 APPROACH

**Data Presentation and Visualization**

The DPV knowledge is used to analyze the huge dataset. The technologies are used are ETL and Visualization. In the dataset, there are some missing variables or incomplete samples which will influence our findings and further predictive modelling.

**ETL**

ETL is the abbreviation of Extract, Transform and Load, which is a technology of extracting data from source databases, transforming and clean the data and load them into a new target database. ETL helps to detect, replace or remove inaccurate, incomplete, irrelevant parts of records from a dataset. It removes inevitable errors and inconsistencies when multiple data sources are pulled together in one dataset. Various tools and programming languages can be used to do the ETL, which improve the efficiency of processing the huge dataset. The most important thing is that ETL can help reflect the intention of data and is easily access to data visualization after ETL process.

**Data Visualization**

Data visualization help people to understand the significance of data, facilitating new discoveries for TCT. Tableau as our visualization tool provides rich diagrams.

**Data Mining**

Data mining the process of identifying patterns and establishing relationships between data and DM currently is widely used in the domain of medical and healthcare. It is used to build predictions of surgery duration in our project.

**Gradient boosting – XGBoost algorithm**

One of the machine learning algorithms: Gradient boosting, which is used for regression problems. The gradient boosting is a method to create a new model, which can predict residuals or errors of previous models, and then add these residuals of errors to make the final prediction. It uses gradient descent algorithms to minimize the loss when adding new models [2]. The pseudocode shows the algorithms of gradient tree boosting (<https://en.wikipedia.org/wiki/Gradient_boosting>). [2]:

XGBoost is an open-source library which provides is an implementation of gradient boosting decision tree. The XGBoost is widely used in data mining.

*Why choose XGBoost?*

Its excellent execution speed and model performance are outstanding. Because it is the top algorithm for competition winners on the Kaggle competitive data science platform [3].

There is also the official XGBoost Python tutorial provided online, which is most efficient to develop a high-performance gradient boosting tree.

Besides, XGBoost enables to lower chance of overfitting.

**K-fold Cross validation**

K-fold cross validation is a way of estimating the predictive performance of our surgical duration model on test data. The motivation to use K-fold cross validation is that the K-fold cross-validation estimator has relatively lower variance than a single hold-out data set estimator. If there are 90% of data are used for training and only 10% used for testing, there will be variation in the performance measures. K-fold validation reduces the variation by dividing into K partitions averagely, so the performance measure becomes less sensitive to the partitioning of the data.

4 EXPERIMENTS

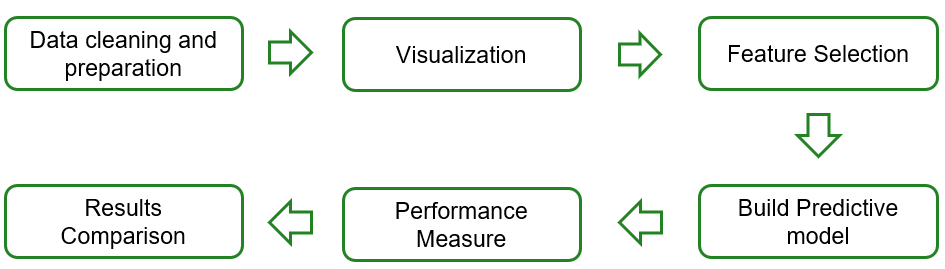


Figure 1 Procedures of experiments

The above is the overall procedures from initials to the predictive model, which consists of six processes.

3.1 Data Cleaning and Preparation

The first step is data cleaning and preparation for the final predictive model. The raw data includes total 4087 surgical cases performed from January 2013 to January 2016 at TCT. This process is to remove the following data:

1. the samples of the missing ratio over 65%
2. the columns of missing ratio over 70%

3.2 Visualizations

The second step is data visualization. The tool we used in this process is Tableau. The key features are extracted and visualized in this process. It provides new discoveries related to surgical duration and the overview of the operation.

3.3 Features Selection

Not all of the data will be used to build the predictive model because some irrelevant data should be deleted, see Table 1.

Table 1. Data Deletion Table

|  |  |
| --- | --- |
| Records | Columns |
| 1. the samples of the missing ratio over 65% 2. the columns of missing ratio over 70% | 1. 2 columns of hospital time (hospital days and IC days)  * (Hospitalization is usually after surgery, so it will not be used in our predictive model. ) |

The remaining data will be as input data for the final predictive model.

3.4 Build Predictive model

Two most popular way of building predictive models: The Gradient boosting decision tree (GBDT) and XGBoost machine learning algorithms are used to predict surgical case duration.

3.5 Performance Measure

The 70% of original set will be hold out as training set and other 30% as final test set.

For the training set, the 5-fold cross validation is used in performance measurement. This process is parallel to predictive model building.

3.6 Result Comparison

In the final step, we will compare the performance with GBDT and XGBoost predictive models separately with planned duration. The standard of judgment is mean squared error (MSE). The smaller MSE means the most accurate performance.

Formula:

5 DISCUSSION

In this section, key findings from main processes will be illustrated.

In the process of data cleaning and preparation, three columns of data are deleted because of data missing according to our previous standards: “Renal\_function”, “Left ventricle”, “Euroscore2”. The evidences can be seen from the below two diagrams.

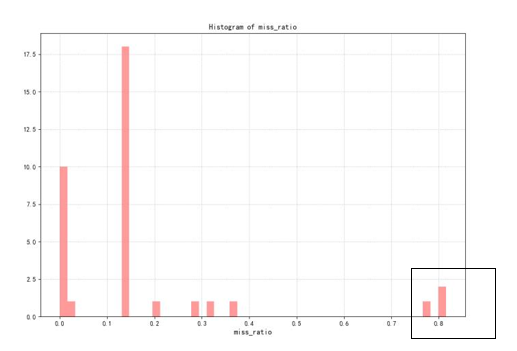


Figure 2 Columns of missing ratio over 70%

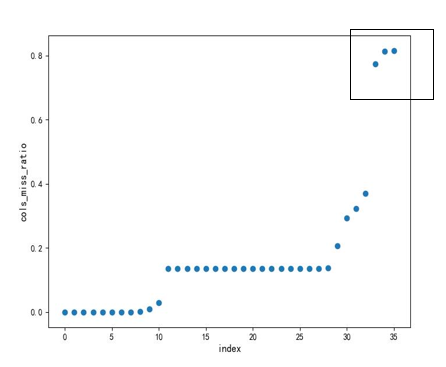


Figure 3 Samples of missing ratio over 65%

After deleting the above three missing data, data visualization is conducted and Tableau diagrams are following.

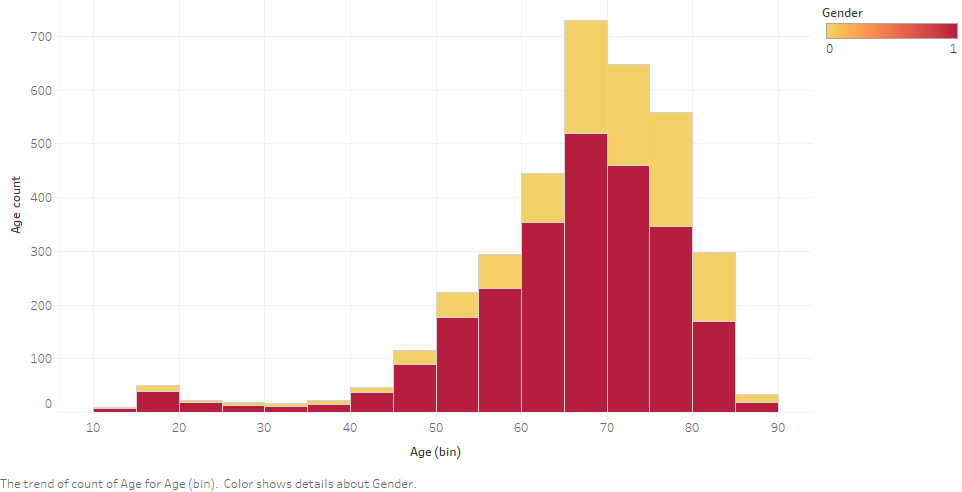


Figure 4 Surgery case patients count

From Fig.4, gender 0 is female and 1 is male, and the X axis represents the value of age range, and the Y axis represents the count of age range, with the age distribution between 60 and 80 years old being the most.

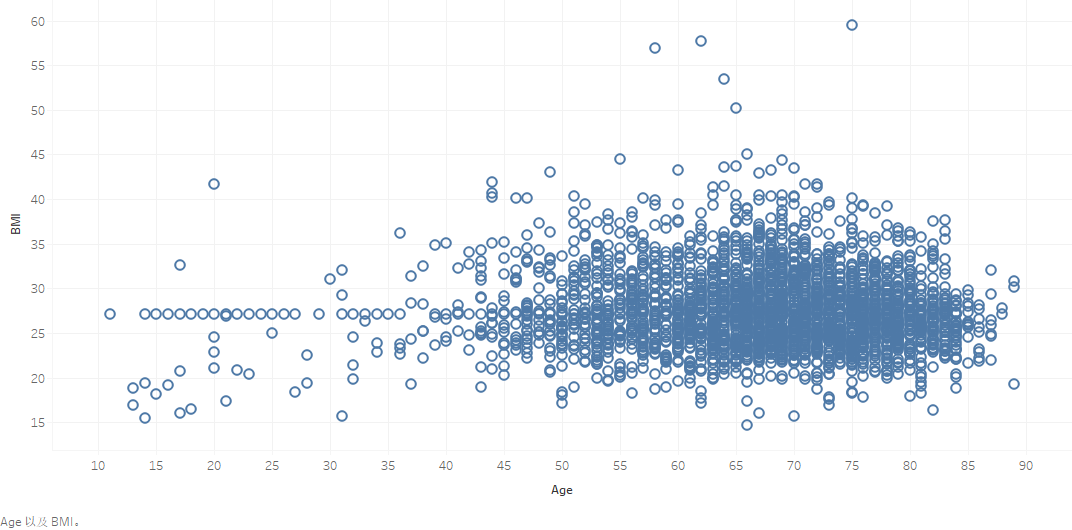


Figure 5 Relation between Patients Age and BMI

The Fig. 5 shows the relation between age of patients and their BMI data.

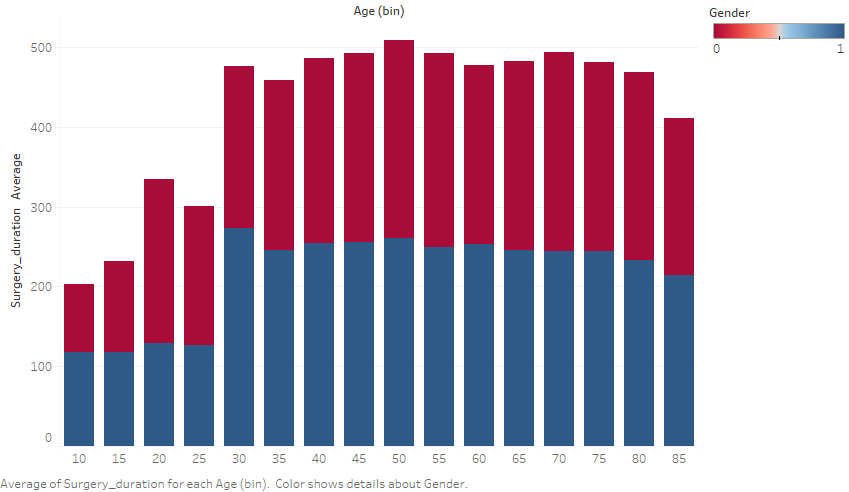


Figure 6 Relation between Age and Average surgery duration

From Fig.4, gender 0 is female and 1 is male, the average surgery duration is different from ages and also gender.

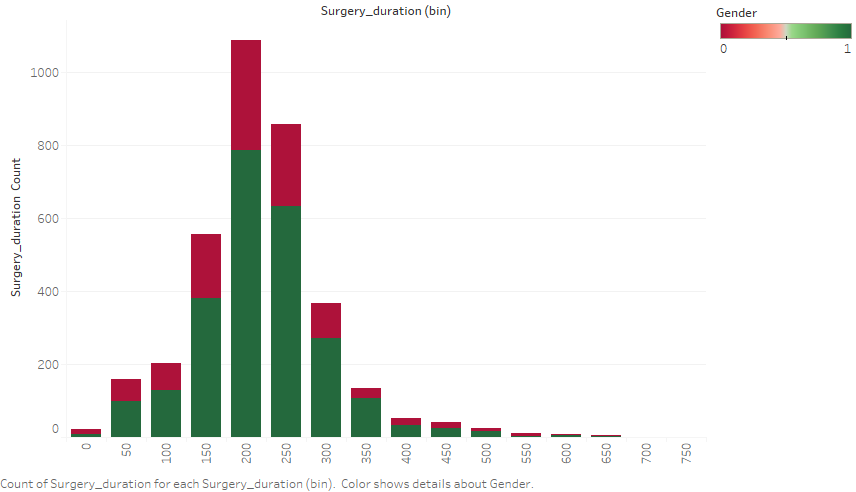


Figure 7 Surgery duration count

From Fig.7, gender 0 is female and 1 is male, the most surgery duration is between the range of 150 to 250 mins. It is not hard to see from the diagram that the surgery duration lasted longer in female patients than in male patients.

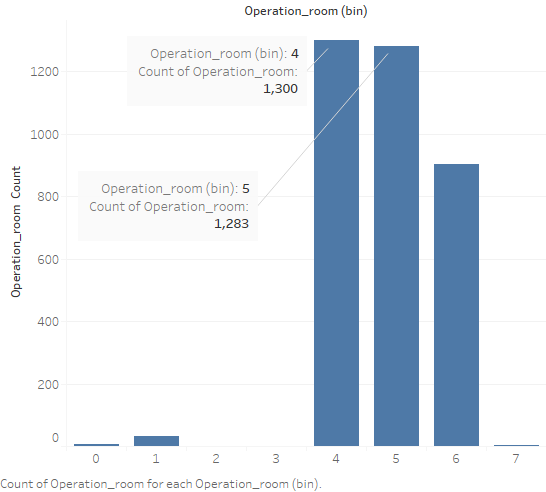


Figure 8 Surgery room and operations

From Fig.8, the number of surgeries performed in each operating room varies, with more than 95% performed in operation room 4, 5, 6.

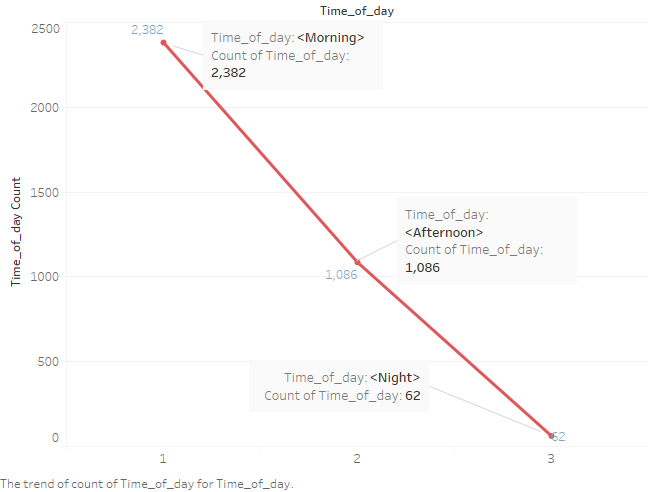


Figure 9 Time of day and operations

From Fig. 9, 1 is morning, 2 is afternoon, 3 is evening, and 2/3 of the surgeries are centered in the morning.

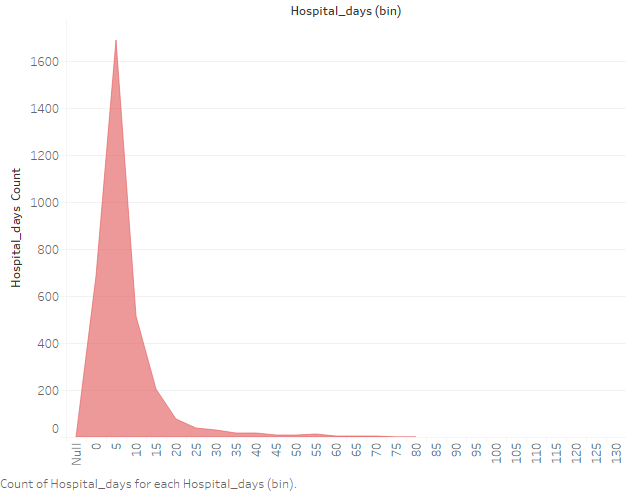


Figure 10 Hospital days counts

From Fig. 10, the days of hospitalization are between 0 and 15 days.

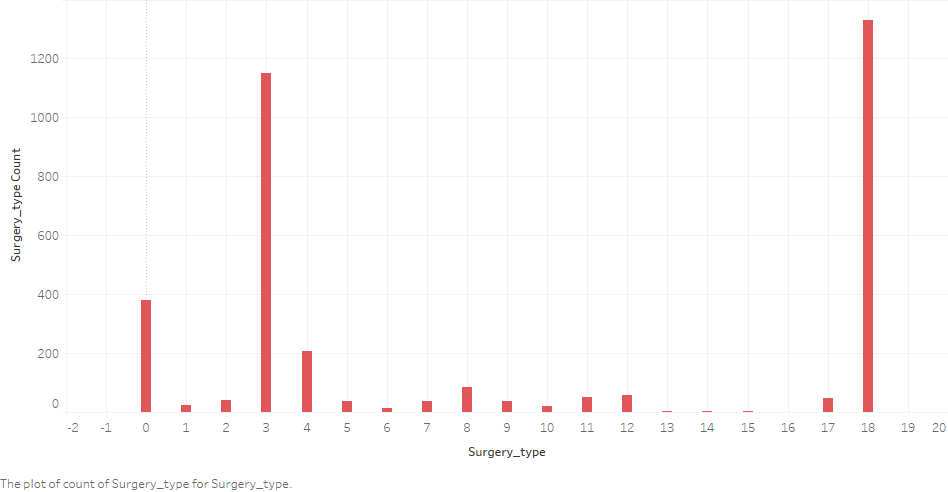


Figure 11 Surgery type counts

From Fig. 11, most surgery types are Coronary Artery Bypass Graft (CABG) and Aortic Valve Replacement (AVR) and other types of surgery type.

Therefore, in the feature selection section, we found that there were total 33 main features (exclude surgery duration and planned surgery duration) in the data set influence the surgical case duration. We deleted 3 missing data after data cleaning and deleted 2 hospital time (hospital days and IC days). Finally remaining 28 main features as the final input for the predictive model. The programming language for building predictive model is Python (see python code in attachment .py files).

In the performance measure step, the overall validation is the following holdout cross validation. The step is 70% as training set and 30% as test set.

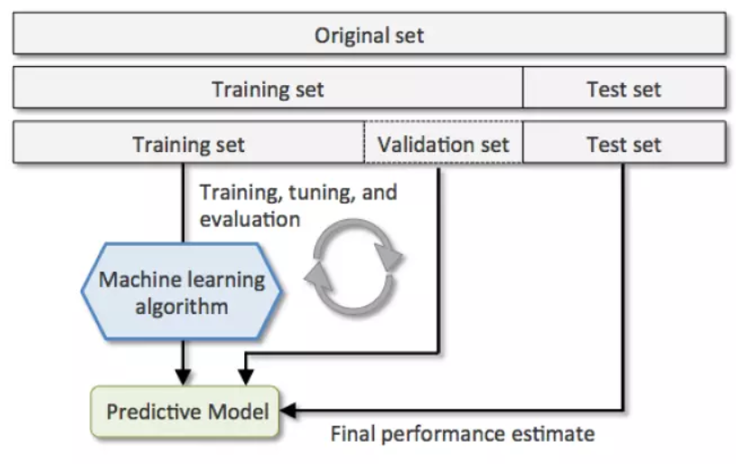


Figure 12 holdout cross validation [4]

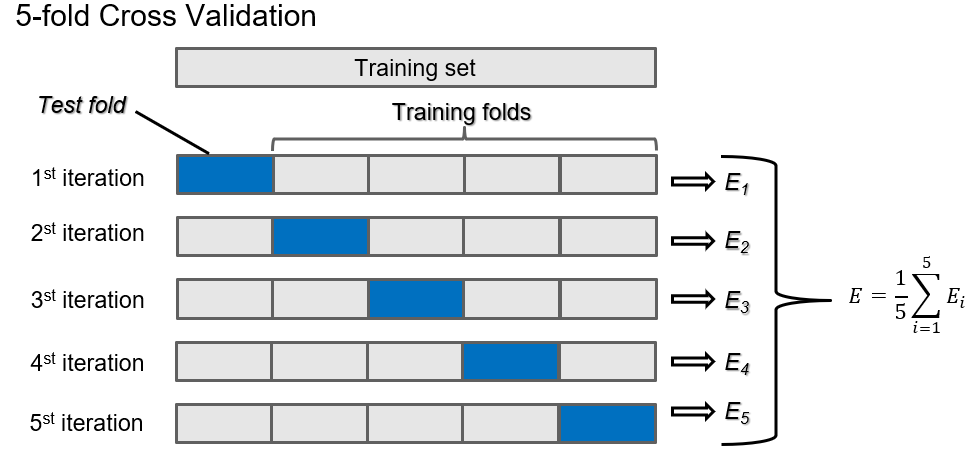


Figure 13 5-fold cross validation

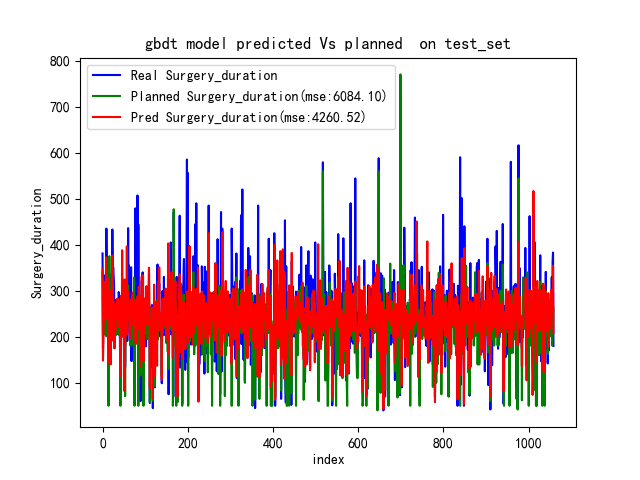


Figure 14 GDBT model predicted VS. planned on test set

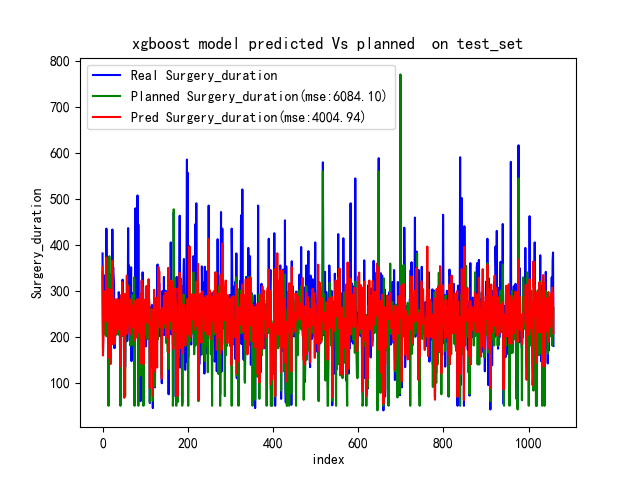


Figure 15 XGBoost model predicted VS. planned on test set

In the final performance measures and results comparison, we got the MSE data of running the GDBT model on test set (4260.52), the MSE data of running the XGBoost model on test set (4004.94), and the MSE data of planned surgery duration from original data running on test set (6084.10).

6 CONCLUSIONS

The conclusion is that XGBoost predictive model has the minimized MSE data, so it is the most accurate predictive model. Besides, from the Tableau, several features related to the patients’ information and operations can be found, which can also provide some evidence for further research.

REFERENCES

[1] Thorax centrum twente, 2017,

<https://www.digitaalpubliceren.com/thoraxcentrumtwente/files/assets/common/downloads/publication.pdf>

[2] <https://en.wikipedia.org/wiki/Gradient_boosting>

[3] Jason Brownlee, 2016, <https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/>

[4]Shan-Huang Wu & Data Lab, 2016, <http://www.cs.nthu.edu.tw/~shwu/courses/ml/labs/08_CV_Ensembling/08_CV_Ensembling.html>