Using the automated machine learning to predict 30-day Hospital Readmission

Tingting Sun   Supervisor: Dr. Lu Xiong

Actuarial Sciences, Department of Mathematical Sciences

Email:ts7f@mtmail.mtsu.edu

The hospital readmission rate is a ratio of the number of people who re-admit 30 days after the last hospital discharge, and it is a metric that both the hospitals and Centers for Medicare & Medicaid Services (CMS) use often. In this research, we used R language to process data and H2O automated machine learning to predict the 30-day hospital readmission rate from 2014 to 2015 for patients older than 65 years. H2O is a machine learning and predictive analytics platform, which is useful, fast, and easy-to-use, allowing non-experts and experts to build machine learning models on large amounts of data. Next, we applied the confusion matrix to determine the False Positive Rate and False Negative Rate in order to compute the accurate rate. We control the coefficients to change the ratio of two different types of error rates. According to our calculations, we improved prediction accuracy , which will help hospitals and CMS reduce the costly readmissions.

What is H2O？

"H2O is an open source, in-memory, distributed, fast, and scalable machine learning and predictive analytics platform that allows you to build machine learning models on big data and provides easy productionalization of those models in an enterprise environment."   -------H2O.ai

Auto machine learning

In recent years, the demand for machine learning experts has outpaced the supply despite the surge of people entering the field. To address this gap, there have been big strides made in the development of user-friendly machine learning software that can be used by non-experts. The first steps toward simplifying machine learning involves developing simple, unified interfaces to a variety of machine learning algorithms.

AutoML (Automatic Machine Learning) is a process that automates the end-to-end process of instantiating real-world problems in machine learning. In a typical machine learning application, practitioners must apply appropriate data preprocessing, feature engineering, and feature extraction. After these pre-steps, practitioners must perform algorithm selection and hyperparameter optimization to convert their final machine learning model to predict performance. Since these steps often exceed the capabilities of non-experts, AutoML is replaced with artificial intelligence-based solutions to meet the growing challenges in machine learning applications. AutoML's end-to-end process provides the advantages of generating simpler solutions that can be created with these solutions and models that are usually manually designed.

Summary data

The raw data contains 17 variables.

AcctNumber	An account number represents a unique admission record
MedicalRecordNumber	The number is used by the hospital as a systematic documentation of a patient´s medical history and care during each hospital stay.
AdmitDate	This is the time when the patient enters the hospital, for some reason.
DischargeDate	This is the time for the patient to be discharged after recovery
AttendingDRCode	Code of attending dcotors
PatientDays	This is the time spent on each admission
AdmitSource	Claim Source Inpatient Admission Code (FFS) The code indicates the source of the referral for the admission .
DischargeStatus	A patient discharge status code is a two-digit code that identifies where the patient is at the conclusion of a health care facility encounter .
DRG	A diagnosis-related group (DRG) is a patient classification system that standardizes prospective payment to hospitals and encourages cost containment initiatives.*(HMSA)
PrimaryProcICD9/PrimaryDiagICD9	ICD-9-CM is the official system of assigning codes to diagnoses and procedures associated with hospital utilization in the United States. The ICD-9 was used to code and classify mortality data from death certificates until 1999, when use of ICD-10 for mortality coding started.*.(wikipedia)
CurrLoc	Current location
MaritalStatus	marital status, are the distinct options that describe a person's relationship with a . , , , and are examples of civil status.*.wikipedia
ZipCode	Zip code
Gender	Male Female
AgeAtDischarge	Age at discharge date

We created a new variable regarding the number of readmissions. When the time interval between the last discharge date and the next admission date is less than thirty days, we label the variable "yes", otherwise, we labeled it "no".This train data contains 3013 records and 1971 patients, while test data contains 607 records. All the patients are 65 years or older.There are 323 readmissions and 2689 non-readmissions.Compared with the admissions marked as "no" with volatility distribution; with the increase of age, the frequency of the admission labeled as "yes" gradually decreases.

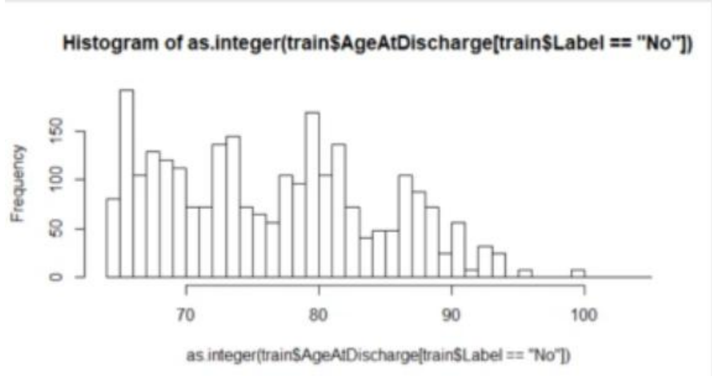

Figure 1:The Frequency of the Admission Labeled as "No"

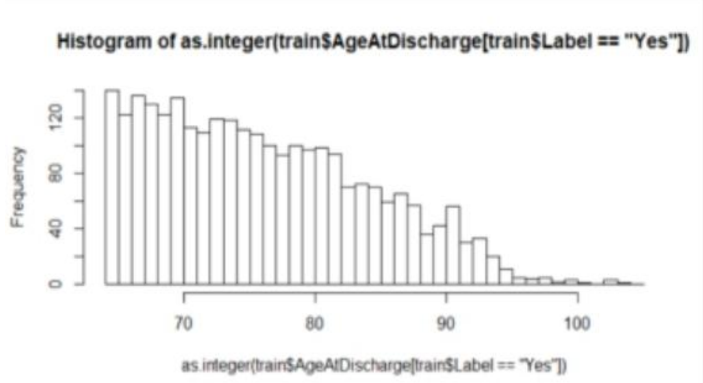

Figure 2:The Frequency of the Admission Labeled as "Yes"

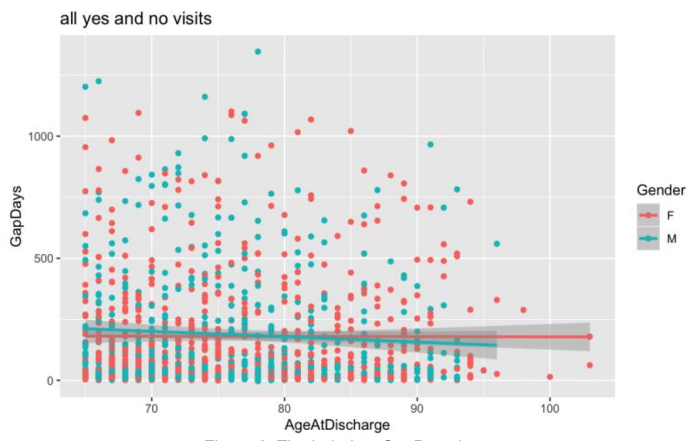

Figure 3: The Label vs GapDays by Gender

We can see for age less than 80 years, females have shorter re-admission days; for age larger than 80 years, males have shorter re-admission days. This is because women usually have a longer lifespan than men. Plot the Label vs GapDays by gender.(Figure 3)

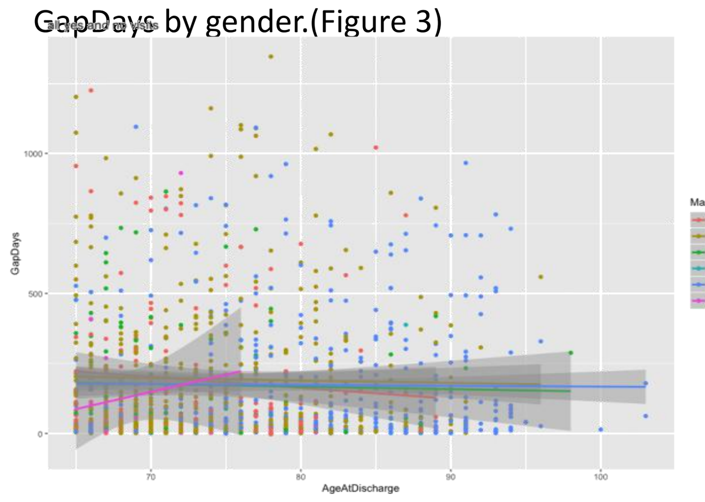

Figure 4: The Label vs GapDays by Marital Status

We can see for age more than 72 years, the separeted have longer readmission days. This means that separedted people need more care and help so that they can spend their later years better. Plot the age vs GapDays by marital status.(Figure 4)

We cannot directly use zipcode as the numerical variable, because zipcode itself has no practical meaning. We need to explore the meaning behind the data. Therefore, we use the average income of different regions to replace zipcode.(Figure 5)

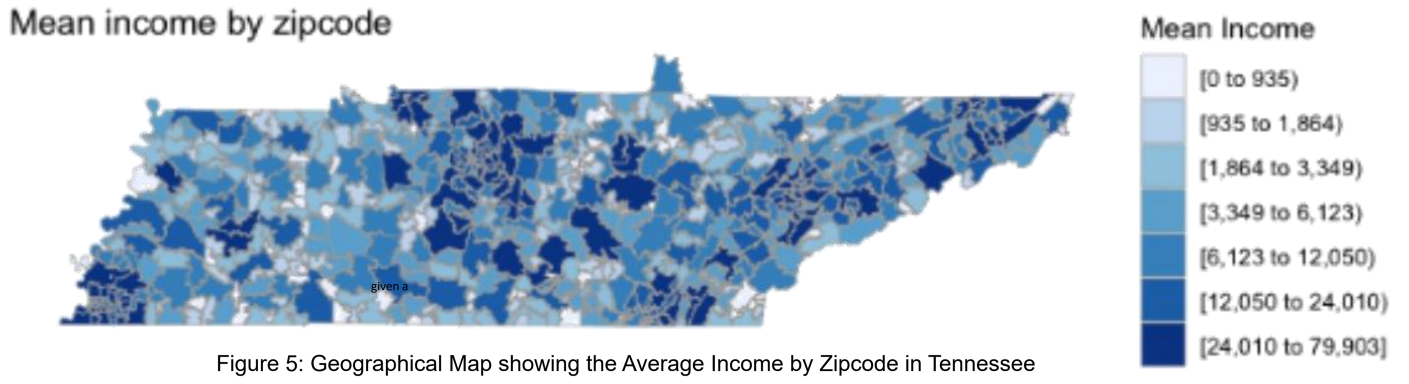

Figure 5: Geographical Map showing the Average Income by Zipcode in Tennessee

optimal model

Through programming, we can find the best model automatically. To prevent overfitting, we used cross validation with n = 5 in the model. There are different indicators to measure the quality of the model, including auc,f1, f2, f0.5 and so on. All the maximum indicators are based on unique(Figure6). The right picture shows details of the optimal model based on auc, which is a common criteria. (Figure 7)

	metric <chr>	threshold <chr>	value <chr>	idx <chr>
1	max f1	0.650749	0.944631	362
2	max f2	0.379394	0.976611	397
3	max f0point5	0.804110	0.927092	263
4	max accuracy	0.661469	0.896117	359
5	max precision	0.993504	1.000000	0
6	max recall	0.379394	1.000000	397
7	max specificity	0.993504	1.000000	0
8	max absolute_mcc	0.842205	0.342941	222
9	max min_per_class_accuracy	0.872246	0.748977	184
10	max mean_per_class_accuracy	0.885844	0.758746	166

Figure 6: The Selected Results by Different Threshold in RStudio

```
H2OBinomialMetrics: glm
** Reported on training data. **

MSE:  0.08258318
RMSE:  0.2873729
LogLoss:  0.2789144
Mean Per-Class Error:  0.4572366
AUC:  0.8304593
AUCPR:  0.9747915
Gini:  0.6609185
R^2:  0.1394937
Residual Deviance:  1680.738
AIC:  4110.738
```

Figure 7:The Detailed Result of the Optimal Model in AUC

AUC (Area under the ROC curve)

This model metric is used to evaluate the degree to which a binary classification model can distinguish between true positives and false positives. AUC of 1 means perfect classifier, and while AUC of 0.5 means poor classifier, its performance is not better than random guessing. H2O uses the trapezoidal rule to approximate the area under the ROC curve. Here is the ROC curve made by different thresholds. What we should do is to choose one. (Figure 8)

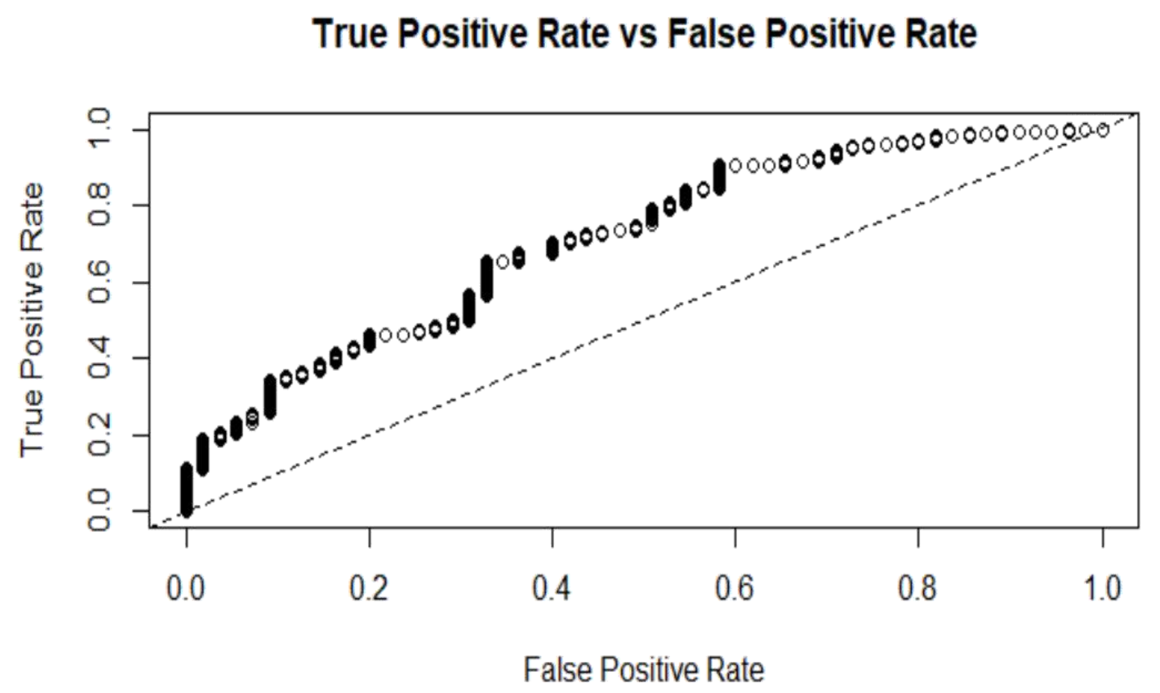

Figure 8: ROC Curve of Train Data

F1, F0.5, F2

- The F1 score is used to measure the ability of the binary classifier to classify positive cases (given a threshold). An F1 score of 1 means that both precision and recall are perfect, and the model can correctly identify all positive cases and will not mark negative cases as positive cases. If the precision or recall rate is low, the F1 score will be close to zero.
- The F0.5 score is the weighted harmonic average of accuracy and recall (given the threshold). The F0.5 scores place greater emphasis on precision and weight than the recall rate. For cases where false positives are considered more serious than false positives, more attention should be paid to accuracy. For example, if your scenario is to predict a product that will run out, you can think that false positives are worse than false positives. In this case, you want your predictions to be very accurate and only capture products that will most likely run out.
- The F2 score is the weighted harmonic average of precision and recall (given threshold). Unlike the F1 score, the F1 score has the same weight precision and recall rate, while the F2 score has a greater weight recall rate (false negative and false positive models have higher penalties). The F2 score is between 0 and 1, where 1 is the ideal model.

formula	weight
$F_{0.5} = 1.25(\frac{(precision)(recall)}{0.25precision+recall})$	precision > recall
$F_1 = 2(\frac{(precision)(recall)}{precision+recall})$	precision = recall
$F_2 = 5(\frac{(precision)(recall)}{4precision+recall})$	precision < recall

Figure 9: The Different Formulas in F0.5, F1, F2

- Precision is the positive observations (true positives) the model correctly identified from all the observations it labeled as positive (the true positives + the false positives).*(wikipedia)
- Pecall is the positive observations (true positives) the model correctly identified from all the actual positive cases (the true positives + the false negatives).*(wikipedia)

Confusion matrix

- The false positive rate is more signficant than true negetive rate.(Figure 10) And for different medical problem, there are different requirements. For example, experts try to decrease the true positive rate in terms of cancer because the consequences of a cancer patient being diagnosed as healthy are more serious than a healthy person being diagnosed with cancer.

	No <chr>	Yes <chr>	Error <chr>	Rate <chr>
No	30	294	0.907407	=294/324
Yes	19	2670	0.007066	=19/2689
Totals	49	2964	0.103883	=313/3013

Figure 10: The Confusion Matrix Based on Auc

In order to satisfy different requirment, there is a simple way to control the balance bewteen the first error rate and the second error rate by controling the threshold. Assume that the cost of the first type of error is ten times that of the second type of error. Index is an indicator of the threshold. Here is line showing the change of total loss. There is a minmun point, which help hospital to decrese penalty.(Figure 11)

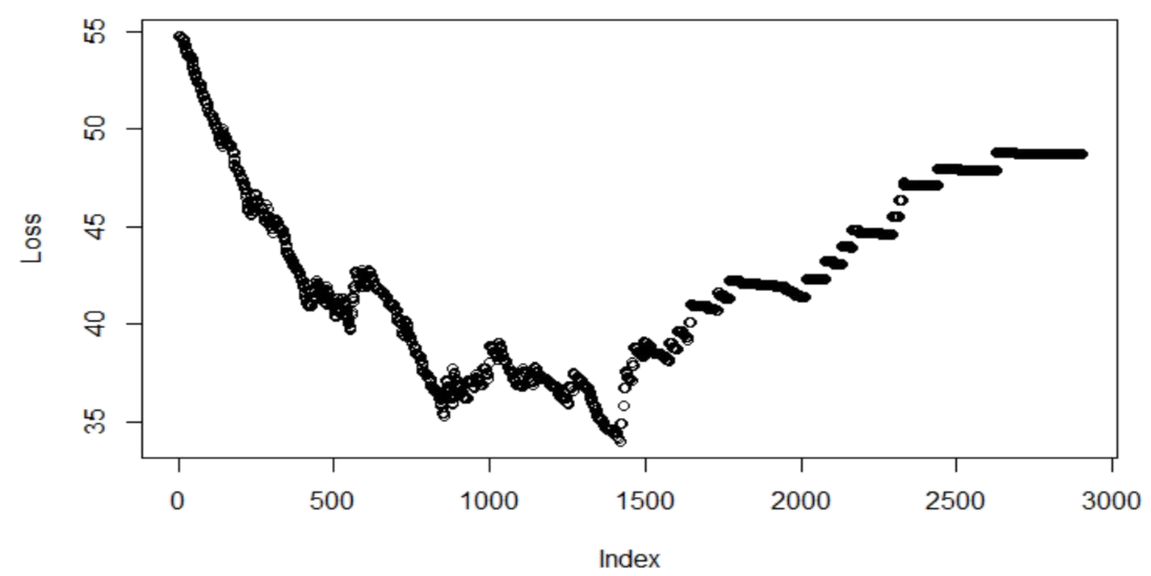

Figure 11: Loss vs Index