DATA698_Master_Thesis_Document

Ali Harb, Dilip Ganesan and Raghunathan Ramnath

4/19/2019

Predict Hospital Readmissions in Diabetes Patients

Abstract:

Today's health care is moving towards value based care. With this proposition in mind, CMS (Center for Medicare & Medicaid Services) came up with the concept of Hospital Readmissions Reduction Program (HRRP). This program is a Medicare value-based purchasing program that reduces payments to hospitals with excess readmissions. The program supports the national goal of improving healthcare for Americans by linking payment to the quality of hospital care. Based on this program the Department of Health and Human Services (HHS) reduce the payments to Inpatient Prospective Payment System (IPPS) hospitals for excess readmissions. Some of the diseases that are classified under the excess readmissions are listed below.

a. Acute Myocardial Infarction (AMI)
b. Chronic Obstructive Pulmonary Disease (COPD)
c. Heart Failure (HF)
d. Pneumonia
e. Coronary Artery Bypass Graft (CABG) Surgery
f. Elective Primary Total Hip Arthroplasty and/or Total Knee Arthroplasty (THA/TKA)

As of FY2018, 6 more new disease conditions have got added to the HRRP program, but as of 2018 the HRRP does not consider diabetes mellitus part of the program. Although diabetes is not yet included in the penalty measures, American hospitals spent over $41 billion on diabetic patients who got readmitted within 30 days of discharge in the year 2011.

Problem Statement:

Predict whether a patient diagnosed with diabetes will be readmitted to hospital within 30 days of discharge.

Research Question:

The focus of this research is to answer the following questions.:

1. Is DRG (Diagnosis Related Group) as parameter a positive predictor for hospital readmission in diabetes mellitus patients?
2. What are the strongest predictors that lead to hospital readmission in diabetes mellitus patients?

Literature Review:

For this project, we have focused our literature review around the parameters that to be used for diabetes hospital readmission prediction. Since our research question is finding the strongest predictor for diabetes hospital readmission, we want to make sure the parameters are qualified for research.

We have grouped the parameters in the data set in to following categories:

Patient Demographics (Age, Race and Sex)

Payment Methodologies (DRG)

Medical Condition and Medications (HbA1C, Insulin and Sulfonylurea et.al)

Let us discuss each of the above in detail below.

Patient Demographics:

While starting this research we strongly believe race and sex as the most important predictor for diabetes. According to [Elias K Spanakis et.al] in the U.S., 8.3% of the population or 25.8 million individuals have diabetes. The prevalence of diabetes is highest among Native Americans (33%) and lowest among Alaska natives (5.5%). Non-Hispanic Whites and Asian Americans have similar prevalence rates of 7.1% and 8.4%, respectively, where Non-Hispanic Blacks and Hispanic Americans overall have higher prevalence rates of 11.8% and 12.6%, respectively. In the article, they went even deeper in their analysis stating, among Hispanic Americans, diabetes varied among their countries of origin. South Americans had one of the lowest prevalence rates (10.1 % in men and 9.8% in women). Similarly, low rates were found among Cuban men and women–13.2% and 13.9%, respectively. The prevalence of diabetes was the highest in those of Mexican, Puerto Rican, Central American, and Dominican descent, with rates of 16.2% to 19.3% for men and 18% to 19.4% for women. This holds good even in Asian American race with Asian Indians have the highest diabetes prevalence whereas Koreans and Japanese have the lowest diabetes rates. Another important parameter along with race is the age of the patient. Per [Elias K Spanakis et.al], The prevalence of diabetes was highest in NHWs in the U.S. between the ages of 0-9 and 10-19. NHB children between the ages of 0-9 and 10-19 years have prevalence, where Hispanic American children have high prevalence's between the ages of 0-9 and 10-19, respectively. All in all, it makes clear that combination of age, race and sex plays a crucial role in diabetes. The above research analysis made us to pick the three parameters for our prediction.

Payment Methodologies:

Though many studies have been done on hospital readmission, the parameters that were used are more clinical or patient centric. Though research were made on the dollar impact because of readmission, very less analysis was done on the parameter of how hospitals were reimbursed. This brought us to the important parameter of DRG (Diagnosis Related Group). Hospital admissions are reimbursed based on DRG. [Joseph Futoma et.al] did the comparison of hospital readmission models based on DRG cohorts. For each visit they have a single Diagnosis Related Group (DRG) code, selected from a set of 815 unique DRGs which break down admissions into broader diagnoses classes than the highly specific ICD codes. They tested a variety of statistical models on 280 different patient-visit cohorts as determined by the DRGs. In the context of regression, this is equivalent to the inclusion of an interaction effect between disease groups and every predictor. According to [M W Rich et. al] there is more financial advantage to hospitals to code patients into more lucrative DRGs, so that patients with more severe disease could conceivably be "promoted" to higher paying DRGs, so the reimbursement increases. From the above research, we feel DRG could be one of the crucial parameter for readmission prediction. Though we do not have the parameter in our dataset, we have got the DRG for the clinical claims data from web scraping.

Medical Condition and Medications:

As far as diabetes is concerned, one of the important parameter is H1A1C test, which measures whether a patient is diabetes or prediabetes or normal. According to [Beata Strack et.al] the decision to obtain a measurement of HbA1c for patients with diabetes mellitus is a valuable predictor for readmission. In their analysis, it showed that the profile of readmission differed significantly in patients where HbA1c was checked in the setting of a primary diabetes diagnosis, when compared to those with a primary circulatory disorder. While readmission rates remained the highest for patients with circulatory diagnoses, readmission rates for patients with diabetes appeared to be associated with the decision to test for HbA1c, rather than the values of the HbA1c result. So, the combination of HbA1c along with primary diagnosis plays are very important role in hospital readmission. Along with other medical condition other important factor is the type of medication which was administered to the patient. According to [Pamela C Heaton et al.] administration of SU[Sulfonylurea] drugs to patient with Type 2 diabetes is associated with an 30% increased risk of readmission compared to other drugs. According to [N. J. Wei] Diabetes medical regimen intensification during hospitalization was not associated with early readmission. Among patients with elevated HbA1c, glucose therapy intensification[Insulin] was associated with a decreased 30-day readmission/emergency department admission risk and lower outpatient HbA1c levels.

Apart from the above researched parameters, we also have other parameters which are available as part of clinical data set.

Modeling:

Apart from the parameters of the dataset we also did some research on the modeling perspective, [Damian Mingle] has done the hospital readmission modeling based on the Extreme Gradient Boosted Tree. Where in the AUC of the machine learning model in greater than that of the LACE score used by the hospitals to determine hospital readmission risks. We will be using the gradient booster for our analysis. Another interesting approach is the use of deep learning to readmission prediction. Per [Ahmad Hammoudeh et.al] the Convolutional neural networks have provided higher AUC compared to other machine learning algorithms. This is another area of interest we are thinking to explore as part of this research thesis.

Data Source:

For this research, we used the Health Facts database (Cerner Corporation, Kansas City, MO), a national data warehouse that collects comprehensive clinical records across hospitals throughout the United States. Health Facts is a voluntary program offered to organizations which use the Cerner Electronic Health Record System. The database contains data systematically collected from participating institutions electronic medical records and includes encounter data (emergency, outpatient, and inpatient), provider specialty, demographics (age, sex, and race), diagnoses and in-hospital procedures documented by ICD-9-CM codes, laboratory data, pharmacy data, in-hospital mortality, and hospital characteristics. All data were de identified in compliance with the Health Insurance Portability and Accountability Act of 1996 before being provided to the investigators. The Health Facts data that is used was an extract representing 10 years (1999–2008) of clinical care at 130 hospitals and integrated delivery networks throughout the United States: Midwest (18 hospitals), Northeast (58), South (28), and West (16). Most of the hospitals (78) have bed size between 100 and 499, 38 hospitals have bed size less than 100, and bed size of 14 hospitals is greater than 500. Because this data represents integrated delivery network health systems in addition to stand-alone hospitals, the data contains both inpatient and outpatient data, including emergency department, for the same group of patients. This dataset is available online at UCI Machine Learning Repository.

Along with the clinical claims data set that is available in UCI, we will also be using the diagnosis cross DRG data set that is available in CMS data source and introduce DRG as a parameter in our model. Deriving of appropriate DRG is complicated because it needs lot more claims information than what is available in the UCI data set, so for our analysis we will be deriving a more appropriate DRG for that episode of care.

The most challenging aspect of data collection is getting the DRG data for this dataset. For extraction of DRG data we went with two approaches.

1. The first approach was to use the CMS crosswalk dataset which is more of an approximate DRG value.
2. The next approach is to send the clincial data with requisite parameter and extract the accurate DRG.

Below you can find a pictorial representation of the process that we have used for data extraction.

We started with the Step 2, the more challenging aspect of data extraction. Normally payer industries use to extract DRG by getting licensed software from different vendors. 3M DRG software is the industry wide solution for getting the DRG. Since using these software costs a lot, we did not want to follow that path. We digged through some research and finally settled on extracting DRG using online Find A Code website.

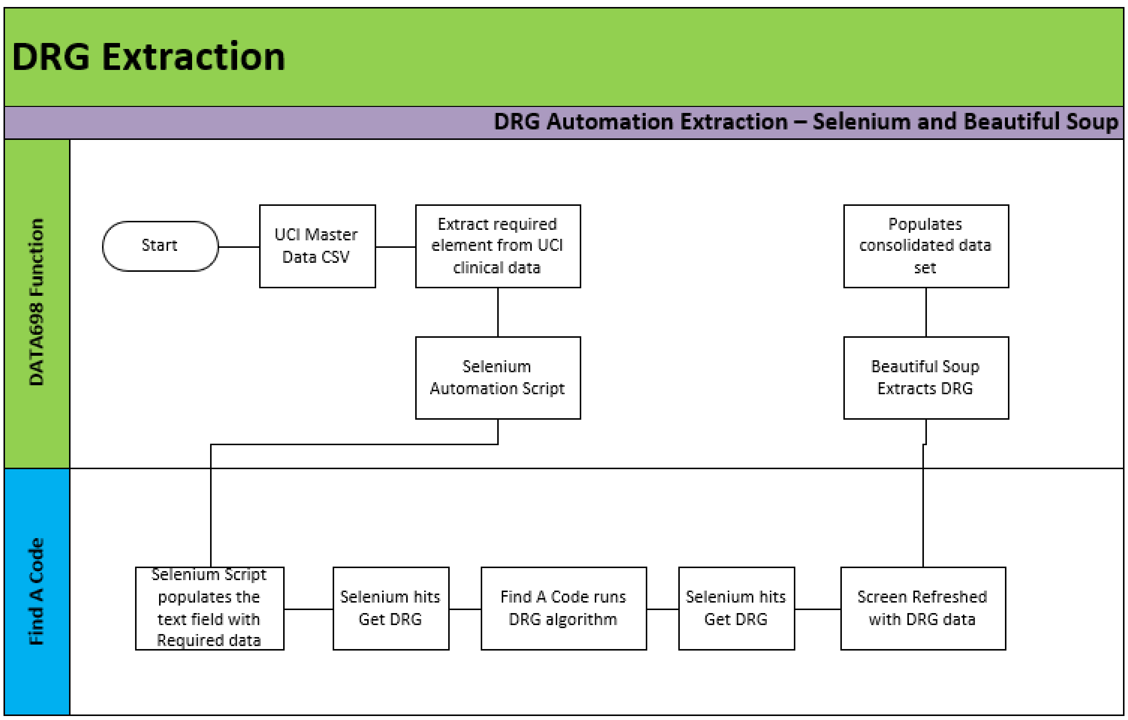


Figure 1: DRG Extract

Find A Code website give DRG when we can send the clinical data online. Some of the clinical data that they require are diagnosis code, patient sex and age. The next problem we stumbled upon is we have 100K rows of record manually submitting the records to the website to extract the DRG was not a feasible task. So we went in to the route of automated extraction of data set. For this automation we relied on Selenium/Python(beautifulsoup) to extract the DRG data.

Selenium is open source automation tool, which helps in submitting the requisite data to a website. We wrote a python code, which will get the input data from xl and go the Find A Code Website and key in the required data in the screen text fields. Once the input data are populated in to those required fields in screen, it automatically goes and click the button get DRG. Once the button is hit on the screen, then the request is send to the server and the Find A Code website does the neccessary algorithm to extract the DRG and displays the same in the screen.

Once the DRG is displayed then we used Python(beautifulsoup) package to go the required field and extract the DRG for that particular input request. Once the DRG was extracted the we wrote the DRG along with the payment for DRG into a output XL file. This process was done in iteration for the entire data set and 100K rows in the data set was filled with DRG and the payment information.

This novel approach not only saved time but also played an important role in the extraction of DRG and in the creation of consolidated data set that is required for this thesis.

Methodology:

The methodology for addressing the problem statement consist of the following processes: Data Exploration, Data Preparation, Regression Modeling, and Ensemble Method modeling with Decision Tree, Random Forest, Neural Network and XG-Boost. The following are discussed below.

Data Exploration:

As a first step in any data analysis, we are going to start with exploration of data set. This step consists of checking whether the variable is numerical or categorical, what is the data type, range of values, plotting to check their distribution, whether are there any NAs and strategy to impute those missing data, how the predictor variables are correlated with one another and with the target variables.

Data Preparation:

As a first step for data preparation, we did lot of data munging, we converted clinical data in to values which could be better used for prediction along with development of strategies for handling missing or invalid data values. Our data set contains a certain degree of class bias and to overcome the bias we developed an approach to the separation of the master data set into dedicated regression modeling “Training” and “Test” subsets.

Logistic Regression:

Identification, development, and testing of task-appropriate regression models. We used a stepwise subtraction method to build the best possible predictive model include log transformations to better fit the data. The confusion matrix was used to select the “best” model which was based on performance metrics including AIC score, AUC, accuracy, classification error rates, precision, specificity, sensitivity, and F1 scores.

Ensemble Method Modeling:

As part of Ensemble methods, we are going to check our prediction using Decision Tree, Random Forrest, Neural Network and XG-Boost algorithms to check whether the accuracy of the model goes up or down.

Results of all models were then compared to see which of these models could better predict hospital readmission and whether DRG has a role in the readmission or not.

Data Exploration:

The diabetes dataset we have used consists of 10,000 records and 54 features. Below given is the detailed description of some of the important features in the dataset.

1. Row ID: Continous running number. Provides a unique identifier for each observation. Missing data 0%
2. Race: Categorical values. Values: African-American Asian Caucasian Hispanic Other. Missing data 2.21%
3. Gender: Categorical values. Values: Male Female. Missing data 0%
4. Age: Categorical values. Values: [0-10) [10-20) [20-30) [30-40) [40-50) [50-60) [60-70) [70-80) [80-90) [90-100). Missing data 0%
5. Weight: Numerical values. Values: Weight of a patient. Missing data 95.92%. Most of the data is missing.
6. Admission type ID: Categorical values. Values: Elective, Emergency, Newborn, Not Available, Not Mapped, Urgent. Missing data 7.21%
7. Discharge disposition ID: Categorical values. Values: 22 levels - Admitted as an inpatient to this hospital, Not Mapped et.al . Missing data 4.69%
8. Admission source ID: Categorical values. Values: 11 levels - Clinic Referral Court/Law, Enforcement Emergency Room HMO Referral Not Available Transfer from another health care facility. Missing data 9.36%
9. Time in hospital: Numerical values. Values: Days spent in the hospital by the patient. Missing data 0%
10. Payer code: Categorical values. Values: BC CH CM CP DM HM MC MD OG OT PO SI SP UN WC. Missing data 53.41%
11. Medical specialty: Categorical values. Values: 52 Levels - Anesthesiology-PediatricCardiology Cardiology-Pediatric Emergency/Trauma . . . Urology. Missing data 41%
12. Number lab procedures: Numerical values. Values: Number of lab procedures done. Missing data 0%
13. Number procedures: Numerical values. Values: Number of procedures done. Missing data 0%
14. Number medications: Numerical values. Values: Number of medications prescribed. Missing data 0%
15. Number outpatient: Numerical values. Values: Number of outpatient visits by the patient. Missing data 0%
16. Number emergency: Numerical values. Values: Number of emergency visits by the patient. Missing data 0%
17. Number inpatient: Numerical values. Values: Number of inpatient visits by the patient. Missing data 0%
18. Diagnoses 1: Categorical values. Values: Different types of diagnoses referred from the ICD-9 codes 457 Levels: 11 110 112 141 150 151 153 154 155 156 157 158 160 161 162 164 171 174 180 182 183 184 185 187 188 . . . V71. Missing data 0.02%
19. Diagnoses 2: Categorical values. Values: Different types of diagnoses referred from the ICD-9 codes 429 Levels: 11 110 112 131 135 136 138 150 151 153 154 155 156 157 162 171 173 174 185 188 189 193 196 197 198 . . . V85. Missing data 0.59%

20. Diagnoses 3: Categorical values. Values: Different types of diagnoses referred from the ICD-9 codes 460 Levels: 110 112 117 135 138 150 151 153 154 155 162 170 172 174 179 185 188 189 196 197 198 199 200 201 202 ... V85. Missing data 2.08%

21. Number diagnoses: Numerical values. Values: Number of diagnoses done by the patient. Missing data 0%

22. Max glu serum: Categorical values. Indicates the range of the result or if the test was not taken. Values: >200, >300, Norm and None if not measured. Missing data 0%

23. A1Cresult: Categorical values. Indicates the range of the result or if the test was not taken. Values: >7 (if result was greater than 7%), >8 (if result was greater than 8%), None (if test not taken), and Norm (if result is normal). Missing data 0%

24. 23 features of medications:

    Metformin, Repaglinide, Nateglinide, Chlorpropamide, Glimepiride, Acetohexamide, Glipizide, Glyburide, Tolbutamide, Pioglitazone, Rosiglitazone, Acarbose, Miglitol, Troglitazone, Tolazamide, Examide, Citoglipton, Insulin, Glyburide metformin, Glipizide metformin, Glimepiride pioglitazone, Metformin rosiglitazone, Metformin pioglitazone. Values: Down (if the dosage is reduce), No (if the medicine is not given), Steady (if the dosage is steady), Up (if the dosage is increased). Missing data 0%

25. Change: Categorical values. Values: Ch (if change in medicine), No (if no change in medicine). Missing data 0%

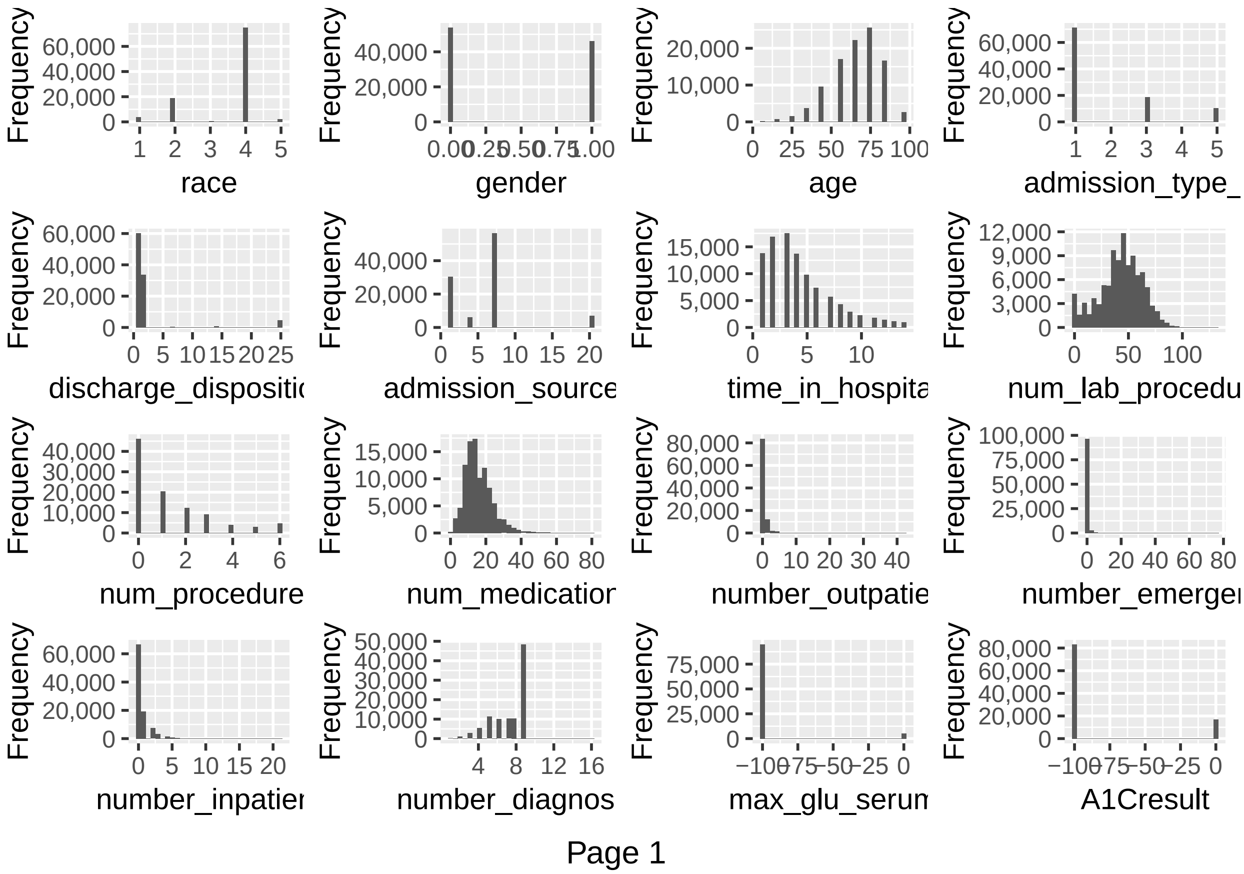
26. Diabetes medicine: Categorical values. Values: No Yes. Missing data 0%

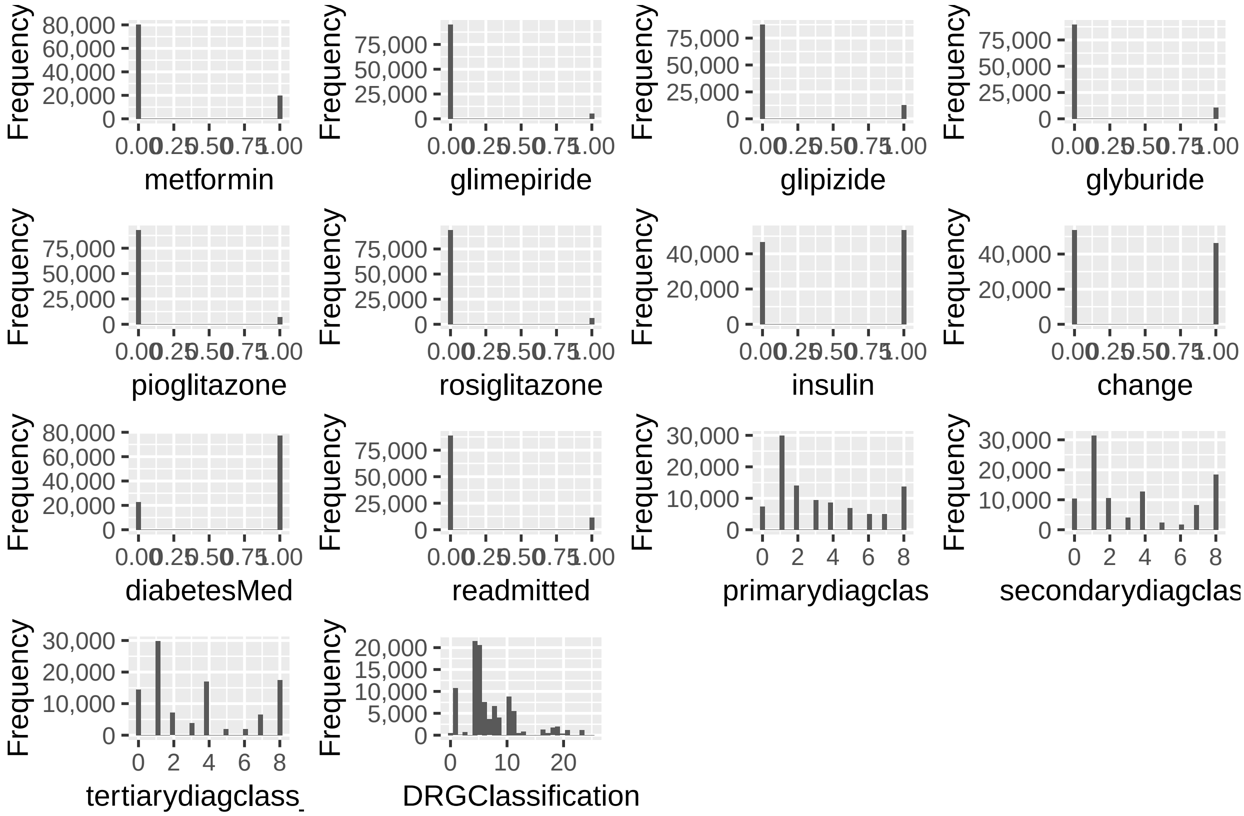
27. Readmitted: Categorical values. Values: FALSE TRUE. Missing data 0%

28. DRG: Categorical values. Values : Different types of Diagnosis Related Group 500 levels: Not many missing values.

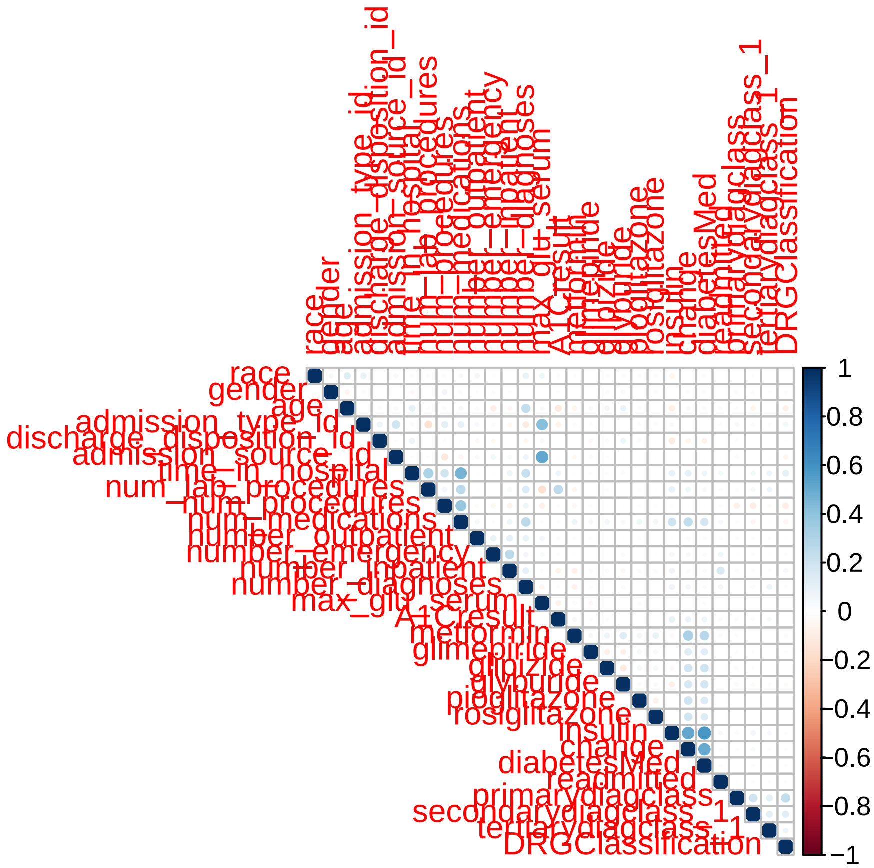
29. DRG Payment: Numerical Values. Values are payment amount for each DRG.

Data Preparation:




Frequency
metformin
glimepiride
glipizide
glyburide
pioglitazone
rosiglitazone
insulin
change
diabetesMed
readmitted
primarydiagclas
secondarydiagclas
tertiarydiagclass_
DRGClassification

Page 2



Training and Test Data for Logistic Regression:

As a first step in data preprocessing, splitting of training and test data set, is to check the class bias. In our dataset there is class bias in our target variable

Ideally, the proportion of events and non-events in the target variable should approximately be the same. So, lets first check the proportion of classes in the dependent variable readmitted.

```
knitr::kable(table(newdataset$readmitted ))
```

Var1	Freq
0	88603
1	11347

```
### Checking of class bias. The number of events happening is less than events not happening. So for pr
```

Clearly, there is a class bias, a condition observed when the proportion of events is much smaller than proportion of non-events. So we must sample the observations in approximately equal proportions to get better models.

As a next step we are going through the process to remove class bias.

One way to address the problem of class bias is to draw the 0s and 1s for the trainingData in equal proportions. In doing so, we will put rest of the inputData not included for training into testData. As a result, the size of trainingData sample will be smaller that validation.

Once the trainingData and testData are created from our dataset, the next step is to create the Binary Logistic Regression.

Logistic Regression:

Binary Regression Base Model

As first step, we are going to run our model using all the variables that are available in the data set. This includes DRGClassification also as predictor variable.

Summary

Analysis of coefficients of Base Model.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.2471730	0.1678836	-7.4287966	0.0000000
race	-0.0165265	0.0181249	-0.9118155	0.3618659
gender	0.0229196	0.0332869	0.6885477	0.4911080
age	0.0062533	0.0011351	5.5088909	0.0000000
admission_type_id	-0.0148371	0.0146160	-1.0151232	0.3100471
discharge_disposition_id	0.0032313	0.0031603	1.0224700	0.3065585
admission_source_id	-0.0098438	0.0042409	-2.3211565	0.0202784
time_in_hospital	0.0188556	0.0066201	2.8482519	0.0043960
num_lab_procedures	0.0008208	0.0009627	0.8525895	0.3938870
num_procedures	-0.0180566	0.0111157	-1.6244183	0.1042866
num_medications	0.0074724	0.0027006	2.7670007	0.0056575
number_outpatient	-0.0140258	0.0125846	-1.1145217	0.2650554
number_emergency	0.0712899	0.0208360	3.4214712	0.0006228
number_inpatient	0.3123913	0.0140707	22.2015559	0.0000000
number_diagnoses	0.0569270	0.0097778	5.8220696	0.0000000
max_glu_serum	0.0026733	0.0009683	2.7609185	0.0057639
A1Cresult	-0.0011097	0.0004718	-2.3521231	0.0186666
metformin	-0.1697504	0.0479709	-3.5386140	0.0004022
glimepiride	-0.1426876	0.0802128	-1.7788630	0.0752622
glipizide	-0.0394206	0.0554962	-0.7103303	0.4774993
glyburide	-0.0770857	0.0612274	-1.2590069	0.2080278
pioglitazone	-0.1051970	0.0673396	-1.5621858	0.1182442
rosiglitazone	-0.0701111	0.0705504	-0.9937734	0.3203332
insulin	-0.0369879	0.0535855	-0.6902590	0.4900313
change	0.0647218	0.0474893	1.3628710	0.1729232
diabetesMed	0.2325196	0.0597013	3.8947149	0.0000983
primarydiagclass	-0.0154751	0.0066070	-2.3422168	0.0191696
secondarydiagclass_1	0.0079339	0.0058183	1.3635954	0.1726950
tertiarydiagclass_1	0.0175397	0.0057634	3.0432966	0.0023400
DRGClassification	0.0044458	0.0036723	1.2106434	0.2260321

The summary(logitMod) gives the beta coefficients, Standard error, z Value and p Value. As a next step of summary analysis we have to look for variables don't turn out to be significant in the model (i.e. p Value turns out greater than significance level of 0.05). The following values are considered to be significant in our model. age, admission_source_id, time_in_hospital, number_emergency, number_inpatient, number_diagnoses, max_glu_serum, A1Cresult, metformin, diabetesMed, primarydiagclass, tertiarydiagclass_1. The above variables becomes the next set of variables for our step wise regression.

Optimal CutOff:

The default cutoff prediction probability score is 0.5 or the ratio of 1's and 0's in the training data. But sometimes, tuning the probability cutoff can improve the accuracy in both the training and test dataset. The optimal cutoff is used to improve the prediction of 1's, 0's, both 1's and 0's and to reduce the misclassification error. Below we will compute the optimal score that we use to minimize the misclassification error for the model. The optimal cutoff value for our model 0.998445

MisClassification Error:

Misclassification error is the percentage mismatch of predcited vs actuals, irrespective of 1's or 0's. The lower the misclassification error, the better is our model. The value for our model is 0.0405

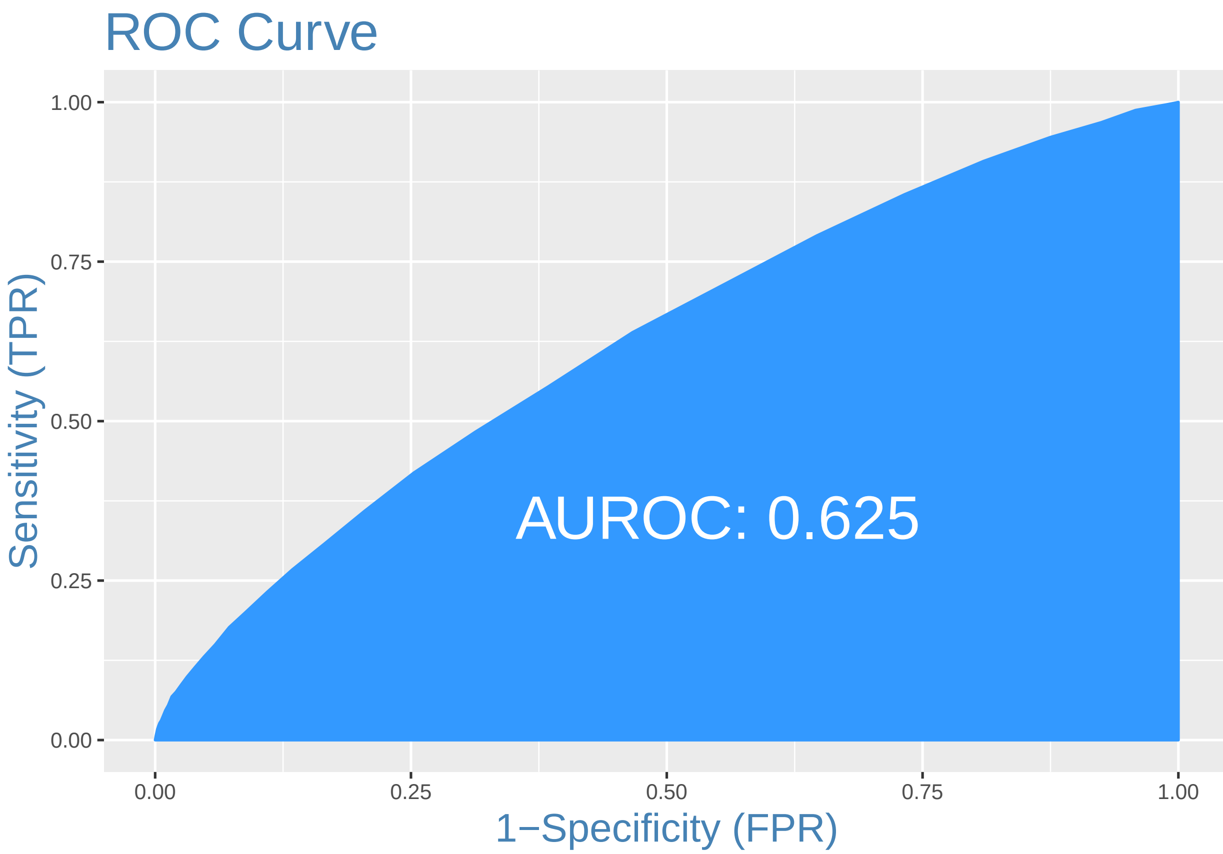
VIF:

From our corrplot analysis we did not find much correlation between our predictor variables and also between predictor variable and target variable. Further as next step in our regression analysis, we want to confirm the same by validating the variance inflation factor. We should check for multicollinearity in the model. As seen below, all predictor variables in the model have VIF well below 4.

	x
race	1.051333
gender	1.020962
age	1.163690
admission_type_id	1.390994
discharge_disposition_id	1.037087
admission_source_id	1.421390
time_in_hospital	1.448778
num_lab_procedures	1.298425
num_procedures	1.290099
num_medications	1.746123
number_outpatient	1.035940
number_emergency	1.096536
number_inpatient	1.111531
number_diagnoses	1.223003
max_glu_serum	1.719808
A1Cresult	1.110510
metformin	1.304300
glimepiride	1.110022
glipizide	1.292438
glyburide	1.297341
pioglitazone	1.109716
rosiglitazone	1.100765
insulin	2.636983
change	2.087960
diabetesMed	2.226846
primarydiagclass	1.117049
secondarydiagclass_1	1.073137
tertiarydiagclass_1	1.035312
DRGClassification	1.115241

ROC:

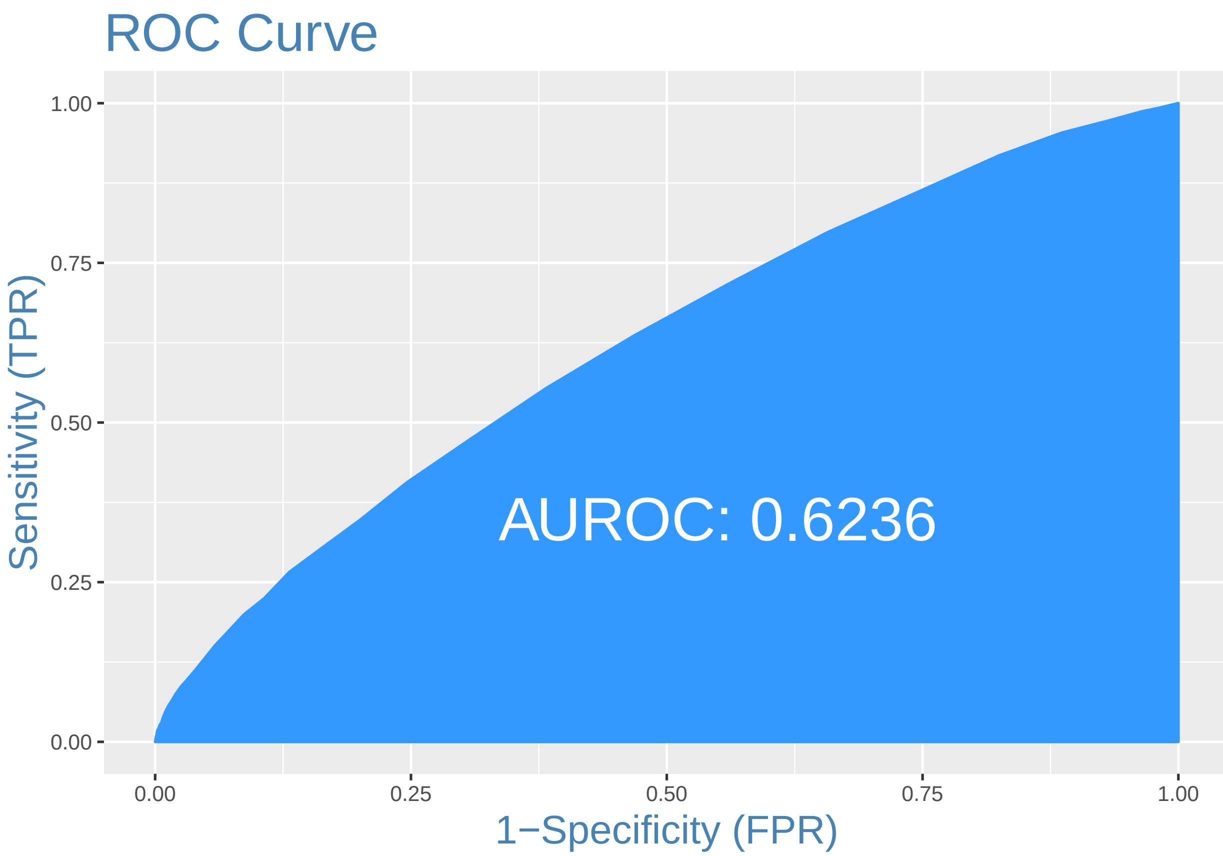
Receiver Operating Characteristics Curve traces the percentage of true positives accurately predicted by a given logit model as the prediction probability cutoff is lowered from 1 to 0. For a good model, as the cutoff is lowered, it should mark more of actual 1's as positives and lesser of actual 0's as 1's. So for a good model, the curve should rise steeply, indicating that the TPR (Y-Axis) increases faster than the FPR (X-Axis) as the cutoff score decreases. Greater the area under the ROC curve, better the predictive ability of the model. We will not look at the curve for our model. From the below curve we can see our curve with AUROC value of 0.625. The value is decent value, though not good.



From the confusion Matrix analysis, we come to the conclusion that our model Accuracy is 67.8752409%. We are able to predict with 67.8752409% accuracy that with DRG as a predictor variable the diabetic mellitus patient will get readmitted. When you look at the Sensitivity of our model it is pretty good. Sensitivity (or True Positive Rate) is the percentage of 1's (actuals) correctly predicted by the model. Which is what we are looking for in our readmission analysis.

Binary Regression Base With Reduced Predictors:

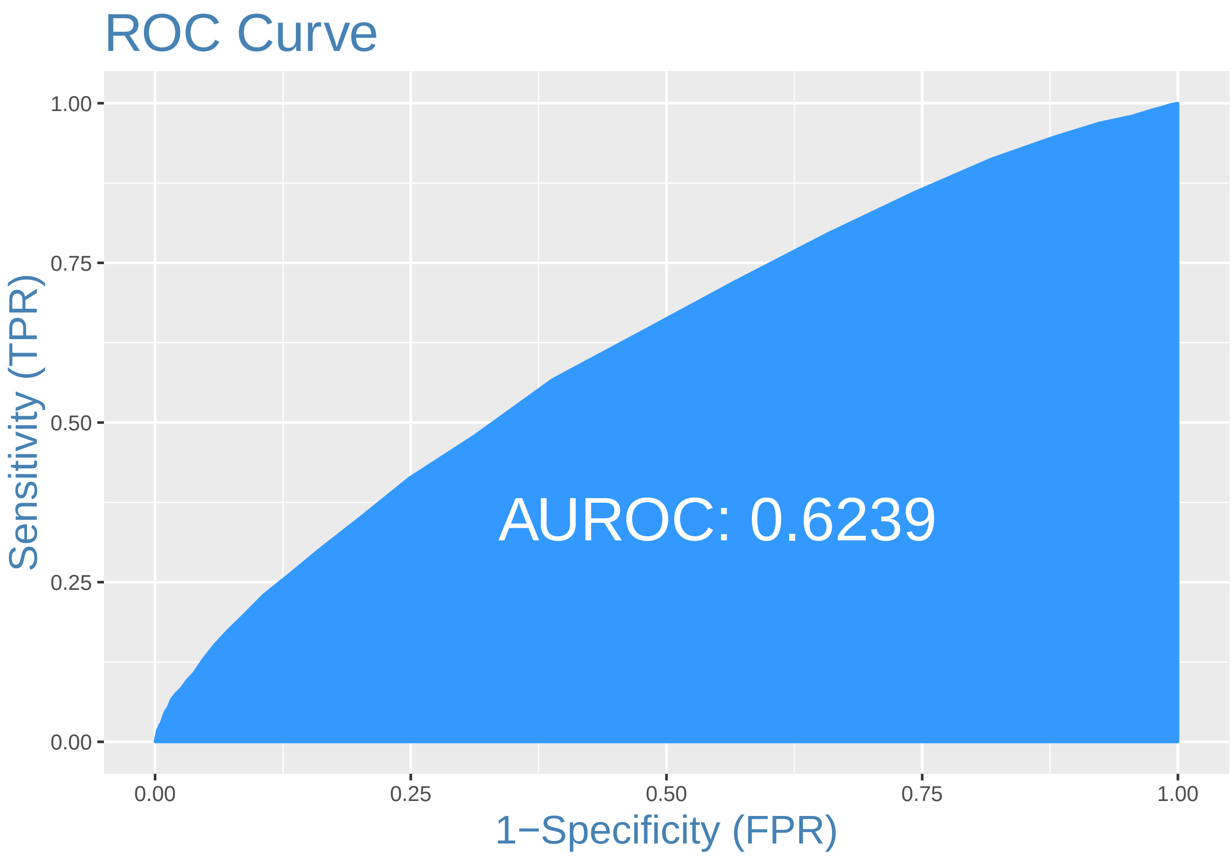
From our first model we are going to drop those predictor variables which we find in statistically less signficant based on p-value(<0.05). With that analysis the list of variables which will be used for this model are. age, admission_source_id, time_in_hospital, number_emergency, number_inpatient, number_diagnoses, max_glu_serum, A1Cresult, metformin, diabetesMed, primarydiagclass, tertiarydiagclass_1



From our reduced predictor variable regression model, the Accuracy of our model has gone up, though not by a greater percent, but to some degree to a value of 68.479528% . This shows that DRGClassification acts a negative parameter from logistic regression modelling prespective. We would like to see how the other logistic regression and ensemble models before drawing conclusions.

Binary Regression Base With Log transformation:

From the plots in our data preparation step, we found some of the variables are skewed either to the left or right. Out of those parameters, the parameters which are important to us as part of our Literature review are Age, Time_In_Hospital and DRGClassification. So in our base model we want to do a log transformation on these parameters and see whether the accuracy our model increases.



For the final model with log transformed variables, the accuracy of the model is 67.9133062%

Conclusion for Binary Logistic Regression:

Comparing the accuracy of the three logistic regression accuracy, the model without DRG had a better accuracy compared to the one with DRG by a smaller margin. So we can conclude that the DRG is not an huge factor in hospital readmission rate for diabetic mellitius patients. Having said that including DRG in the model does not decreases the accuracy by greater margin. So it is more of an neutral effect. Below is the accuracy chart of three models.

Model	Accurarcy	ClassErrorRate	Precision	Senstivity	Specificity	F1
Model with DRG	67.87524	0.3212476	0.6870111	0.9692173	0.0611729	0.8040715
Model No DRG	68.47953	0.3152047	0.6937801	0.9688707	0.0610864	0.8085681
Model with Log Transformed Variables	67.91331	0.3208669	0.6875690	0.9690214	0.0608206	0.8043860