azizhazeinita

```
#1. Use same train and test sample from logistic regression (assignment 4.1).

df <- read.csv("/Users/azizhazeinita/Documents/S2 Uchicago/MScA/Winter 2022 - 2nd Quarter/Data Mining Principles/Assignment 4 - Data Mining Principles/Diabetese Dataset Files/diabetes_data.csv")
df_1 <- read.csv("/Users/azizhazeinita/Documents/S2 Uchicago/MScA/Winter 2022 - 2nd Quarter/Data Mining Principles/Assignment 4 - Data Mining Principles/Diabetese Dataset Files/diabetes_data.csv")
colSums(is.na(df))
```

```
##              encounter_id               patient_nbr                      race
##                         0                         0                         0
##                    gender                       age        admission_type_id
##                         0                         0                         0
## discharge_disposition_id      admission_source_id         time_in_hospital
##                         0                         0                         0
##       num_lab_procedures          num_procedures          num_medications
##                         0                         0                         0
##        number_outpatient         number_emergency         number_inpatient
##                         0                         0                         0
##                    diag_1                    diag_2                    diag_3
##                         0                         0                         0
##         number_diagnoses            max_glu_serum                A1Cresult
##                         0                         0                         0
##                 metformin             repaglinide             nateglinide
##                         0                         0                         0
##           chlorpropamide             glimepiride            acetohexamide
##                         0                         0                         0
##                 glipizide               glyburide             tolbutamide
##                         0                         0                         0
##             pioglitazone           rosiglitazone                 acarbose
##                         0                         0                         0
##                 miglitol            troglitazone               tolazamide
##                         0                         0                         0
##                  insulin      glyburide.metformin      glipizide.metformin
##                         0                         0                         0
## glimepiride.pioglitazone   metformin.pioglitazone                    change
##                         0                         0                         0
##              diabetesMed               readmitted
##                         0                         0
```

```
df <- na.omit(df)
colSums(is.na(df_1))
```

```
##              encounter_id               patient_nbr                      race
##                         0                         0                         0
##                    gender                       age         admission_type_id
##                         0                         0                         0
## discharge_disposition_id       admission_source_id         time_in_hospital
##                         0                         0                         0
##        num_lab_procedures           num_procedures          num_medications
##                         0                         0                         0
##         number_outpatient         number_emergency         number_inpatient
##                         0                         0                         0
##                    diag_1                    diag_2                    diag_3
##                         0                         0                         0
##          number_diagnoses            max_glu_serum                A1Cresult
##                         0                         0                         0
##                 metformin             repaglinide              nateglinide
##                         0                         0                         0
##            chlorpropamide             glimepiride            acetohexamide
##                         0                         0                         0
##                 glipizide               glyburide              tolbutamide
##                         0                         0                         0
##              pioglitazone           rosiglitazone                  acarbose
##                         0                         0                         0
##                 miglitol            troglitazone               tolazamide
##                         0                         0                         0
##                   insulin      glyburide.metformin      glipizide.metformin
##                         0                         0                         0
## glimepiride.pioglitazone   metformin.pioglitazone                    change
##                         0                         0                         0
##               diabetesMed               readmitted
##                         0                         0
```

```
df <- na.omit(df_1)

df_readmitted <- df_1['readmitted']
df_readmitted['readmitted'][df_readmitted['readmitted'] == '<30'] <- 'Yes'
df_readmitted['readmitted'][df_readmitted['readmitted'] == '>30'] <- 'Yes'

# Changing diag_1, diag_2, diag_3 type into float
df_diag <- df_1[c('diag_1','diag_2','diag_3')]
str(df_diag)
```

```
## 'data.frame':    98053 obs. of  3 variables:
##  $ diag_1: chr  "276" "648" "8" "197" ...
##  $ diag_2: chr  "250.01" "250" "250.43" "157" ...
##  $ diag_3: chr  "255" "V27" "403" "250" ...
```

```
sum(is.na(df_diag))
```

```
## [1] 0
```

```
df_diag$diag_1 <- gsub("[^0-9.-]", "", df_diag$diag_1)
df_diag$diag_1 <- as.double(df_diag$diag_1)
df_diag$diag_2 <- gsub("[^0-9.-]", "", df_diag$diag_2)
df_diag$diag_2 <- as.double(df_diag$diag_2)
df_diag$diag_3 <- gsub("[^0-9.-]", "", df_diag$diag_3)
df_diag$diag_3 <- as.double(df_diag$diag_3)

df[c('readmitted','diag_1','diag_2','diag_3')] <- NULL

library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
dummy <- dummyVars(" ~ .", data=df) #define one-hot encoding function
final_df <- data.frame(predict(dummy, newdata=df)) #perform one-hot encoding on data frame

unique(df_readmitted)
```

```
##   readmitted
## 1        Yes
## 2         NO
```

```
df_readmitted$readmitted = factor(df_readmitted$readmitted, levels = c('Yes', 'NO'),labels
= c(1, 0))

final_df['readmitted'] <- df_readmitted
final_df['diag_1'] <- df_diag$diag_1
final_df['diag_2'] <- df_diag$diag_2
final_df['diag_3'] <- df_diag$diag_3

str(final_df)
```

```
## 'data.frame':    98053 obs. of  114 variables:
##  $ encounter_id                : num  149190 64410 500364 16680 35754 ...
##  $ patient_nbr                 : num  55629189 86047875 82442376 42519267 82637451
...
##  $ raceAfricanAmerican         : num  0 1 0 0 0 0 0 0 0 1 ...
##  $ raceAsian                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ raceCaucasian               : num  1 0 1 1 1 1 1 1 1 0 ...
##  $ raceHispanic                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ raceOther                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ genderFemale                : num  1 1 0 0 0 0 0 1 1 1 ...
##  $ genderMale                  : num  0 0 1 1 1 1 1 0 0 0 ...
##  $ genderUnknown.Invalid       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ age.0.10.                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ age.10.20.                  : num  1 0 0 0 0 0 0 0 0 0 ...
##  $ age.20.30.                  : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ age.30.40.                  : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ age.40.50.                  : num  0 0 0 1 0 0 0 0 0 1 ...
##  $ age.50.60.                  : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ age.60.70.                  : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ age.70.80.                  : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ age.80.90.                  : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ age.90.100.                 : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ admission_type_id           : num  1 1 1 1 2 3 1 2 3 1 ...
##  $ discharge_disposition_id    : num  1 1 1 1 1 1 1 1 3 1 ...
##  $ admission_source_id         : num  7 7 7 7 2 2 7 4 4 7 ...
##  $ time_in_hospital            : num  3 2 2 1 3 4 5 13 12 9 ...
##  $ num_lab_procedures          : num  59 11 44 51 31 70 73 68 33 47 ...
##  $ num_procedures              : num  0 5 1 0 6 1 0 2 3 2 ...
##  $ num_medications             : num  18 13 16 8 16 21 12 28 18 17 ...
##  $ number_outpatient           : num  0 2 0 0 0 0 0 0 0 0 ...
##  $ number_emergency            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ number_inpatient            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ number_diagnoses            : num  9 6 7 5 9 7 8 8 8 9 ...
##  $ max_glu_serum.200           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ max_glu_serum.300           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ max_glu_serumNone           : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ max_glu_serumNorm           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ A1Cresult.7                 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ A1Cresult.8                 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ A1CresultNone               : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ A1CresultNorm               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ metforminDown               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ metforminNo                 : num  1 1 1 1 1 0 1 1 1 1 ...
##  $ metforminSteady             : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ metforminUp                 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ repaglinideDown             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ repaglinideNo               : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ repaglinideSteady           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ repaglinideUp               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ nateglinideDown             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ nateglinideNo               : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ nateglinideSteady           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ nateglinideUp               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ chlorpropamideDown          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ chlorpropamideNo            : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ chlorpropamideSteady        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ chlorpropamideUp            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ glimepirideDown             : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ glimepirideNo                : num  1 1 1 1 1 0 1 1 1 1 ...
##  $ glimepirideSteady            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ glimepirideUp                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acetohexamideNo              : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ acetohexamideSteady          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ glipizideDown                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ glipizideNo                  : num  1 0 1 0 1 1 1 0 1 1 ...
##  $ glipizideSteady              : num  0 1 0 1 0 0 0 1 0 0 ...
##  $ glipizideUp                  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ glyburideDown                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ glyburideNo                  : num  1 1 1 1 1 1 0 1 1 1 ...
##  $ glyburideSteady              : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ glyburideUp                  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ tolbutamideNo                : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ tolbutamideSteady            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ pioglitazoneDown             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ pioglitazoneNo               : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ pioglitazoneSteady           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ pioglitazoneUp               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ rosiglitazoneDown            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ rosiglitazoneNo              : num  1 1 1 1 1 1 1 1 0 1 ...
##  $ rosiglitazoneSteady          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ rosiglitazoneUp              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acarboseDown                 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acarboseNo                   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ acarboseSteady               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acarboseUp                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ miglitolDown                 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ miglitolNo                   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ miglitolSteady               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ miglitolUp                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ troglitazoneNo               : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ troglitazoneSteady           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ tolazamideNo                 : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ tolazamideSteady             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ tolazamideUp                 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ insulinDown                  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ insulinNo                    : num  0 1 0 0 0 0 1 0 0 0 ...
##  $ insulinSteady                : num  0 0 0 1 1 1 0 1 1 1 ...
##  $ insulinUp                    : num  1 0 1 0 0 0 0 0 0 0 ...
##  $ glyburide.metforminDown      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ glyburide.metforminNo        : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ glyburide.metforminSteady    : num  0 0 0 0 0 0 0 0 0 0 ...
##   [list output truncated]
```

```
smp_size <- floor(0.7 * nrow(final_df)) #70% data train
set.seed(123) # set the seed to make your partition reproducible
train_ind <- sample(seq_len(nrow(final_df)), size = smp_size)

train <- final_df[train_ind, ]
test <- final_df[-train_ind, ]
str(train)
```

```
## 'data.frame':    68637 obs. of  114 variables:
##  $ encounter_id                  : num  1.58e+08 1.69e+08 2.00e+07 1.01e+08 4.06e+08
...
##  $ patient_nbr                   : num  1.12e+08 6.26e+07 2.36e+07 2.42e+07 1.81e+08
...
##  $ raceAfricanAmerican           : num  0 0 0 1 0 0 0 1 0 0 ...
##  $ raceAsian                     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ raceCaucasian                 : num  1 1 1 0 0 1 1 0 1 1 ...
##  $ raceHispanic                  : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ raceOther                     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ genderFemale                  : num  0 1 0 1 1 1 1 1 0 1 ...
##  $ genderMale                    : num  1 0 1 0 0 0 0 0 1 0 ...
##  $ genderUnknown.Invalid         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ age.0.10.                     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ age.10.20.                    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ age.20.30.                    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ age.30.40.                    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ age.40.50.                    : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ age.50.60.                    : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ age.60.70.                    : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ age.70.80.                    : num  1 1 0 1 0 0 0 0 0 0 ...
##  $ age.80.90.                    : num  0 0 0 0 0 0 1 0 0 1 ...
##  $ age.90.100.                   : num  0 0 1 0 0 0 0 1 0 0 ...
##  $ admission_type_id             : num  2 2 5 1 1 1 1 1 5 1 ...
##  $ discharge_disposition_id      : num  1 1 11 3 1 1 3 6 6 1 ...
##  $ admission_source_id           : num  7 1 17 7 7 7 7 7 1 7 ...
##  $ time_in_hospital              : num  4 1 1 6 9 4 7 12 5 5 ...
##  $ num_lab_procedures            : num  65 9 17 45 75 66 57 74 71 63 ...
##  $ num_procedures                : num  1 0 0 1 2 0 0 1 1 1 ...
##  $ num_medications               : num  19 5 7 14 17 11 9 15 32 22 ...
##  $ number_outpatient             : num  0 0 0 0 0 2 0 0 2 0 ...
##  $ number_emergency              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ number_inpatient              : num  0 0 0 0 0 8 0 0 1 4 ...
##  $ number_diagnoses              : num  9 9 8 9 9 7 9 9 9 9 ...
##  $ max_glu_serum.200             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ max_glu_serum.300             : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ max_glu_serumNone             : num  1 1 0 1 1 1 1 1 1 1 ...
##  $ max_glu_serumNorm             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ A1Cresult.7                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ A1Cresult.8                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ A1CresultNone                 : num  0 1 1 1 0 1 0 1 1 1 ...
##  $ A1CresultNorm                 : num  1 0 0 0 1 0 1 0 0 0 ...
##  $ metforminDown                 : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ metforminNo                   : num  1 1 1 1 1 1 0 1 0 1 ...
##  $ metforminSteady               : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ metforminUp                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ repaglinideDown               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ repaglinideNo                 : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ repaglinideSteady             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ repaglinideUp                 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ nateglinideDown               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ nateglinideNo                 : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ nateglinideSteady             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ nateglinideUp                 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ chlorpropamideDown            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ chlorpropamideNo              : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ chlorpropamideSteady          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ chlorpropamideUp              : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ glimepirideDown               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ glimepirideNo                 : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ glimepirideSteady             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ glimepirideUp                 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acetohexamideNo               : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ acetohexamideSteady           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ glipizideDown                 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ glipizideNo                   : num  1 1 1 1 0 1 1 1 1 1 ...
##  $ glipizideSteady               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ glipizideUp                   : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ glyburideDown                 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ glyburideNo                   : num  1 1 1 1 1 1 1 1 0 0 ...
##  $ glyburideSteady               : num  0 0 0 0 0 0 0 0 1 1 ...
##  $ glyburideUp                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ tolbutamideNo                 : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ tolbutamideSteady             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ pioglitazoneDown              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ pioglitazoneNo                : num  1 1 1 1 1 1 1 1 1 0 ...
##  $ pioglitazoneSteady            : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ pioglitazoneUp                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ rosiglitazoneDown             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ rosiglitazoneNo               : num  1 1 1 1 1 1 0 1 1 1 ...
##  $ rosiglitazoneSteady           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ rosiglitazoneUp               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acarboseDown                  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acarboseNo                    : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ acarboseSteady                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ acarboseUp                    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ miglitolDown                  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ miglitolNo                    : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ miglitolSteady                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ miglitolUp                    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ troglitazoneNo                : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ troglitazoneSteady            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ tolazamideNo                  : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ tolazamideSteady              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ tolazamideUp                  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ insulinDown                   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ insulinNo                     : num  0 1 0 0 0 1 1 1 0 0 ...
##  $ insulinSteady                 : num  0 0 1 1 1 0 0 0 1 0 ...
##  $ insulinUp                     : num  1 0 0 0 0 0 0 0 0 1 ...
##  $ glyburide.metforminDown       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ glyburide.metforminNo         : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ glyburide.metforminSteady     : num  0 0 0 0 0 0 0 0 0 0 ...
##   [list output truncated]
```

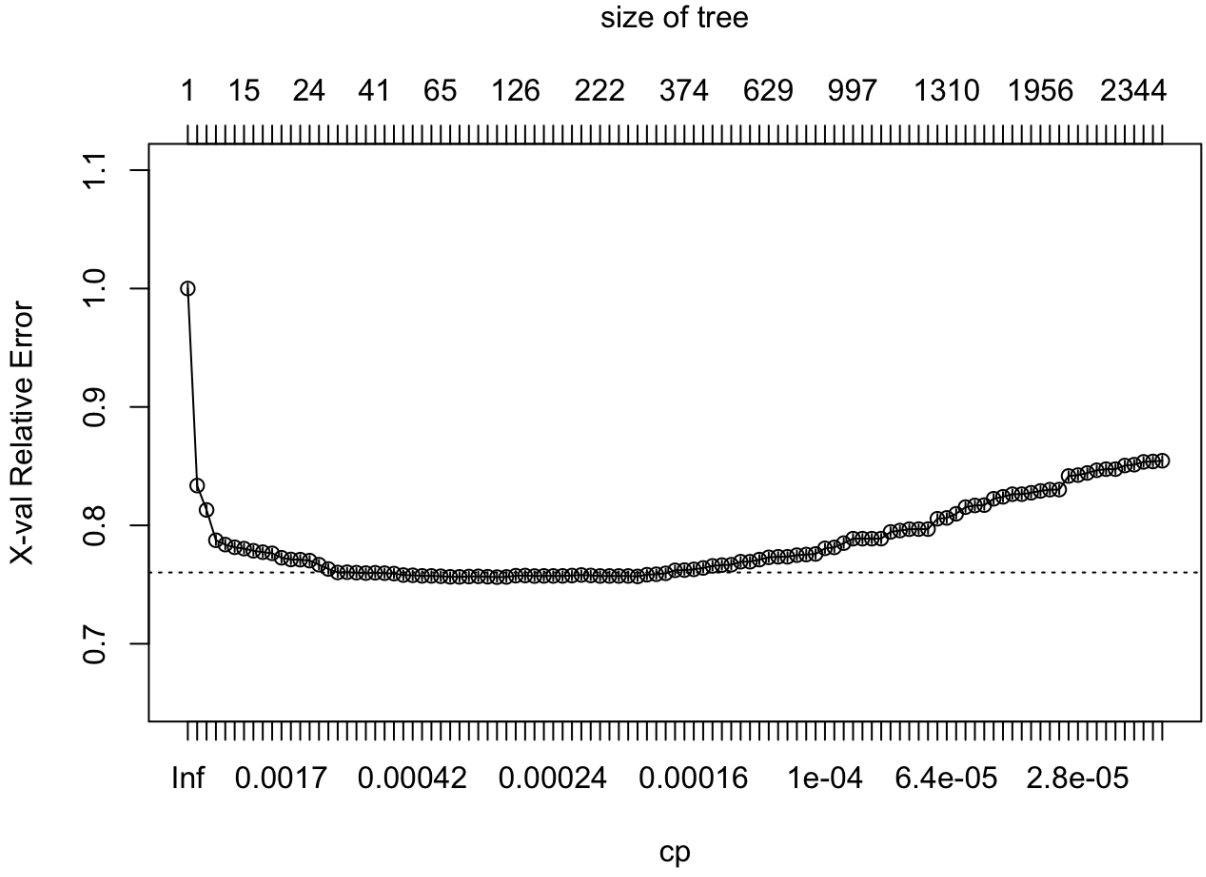
```
# 2. Build a full classification tree using the train data and all variables.

library(rpart)

# 2.1 Build the model with criteria: minimum node size = 30, Use 10-fold cross validation
 (Xval = 10), CP should = 0
x=rpart(readmitted~., data=train,control=rpart.control(cp=0,minsplit=30,xval=10))

# 2.2. View results for different tree lengths
printcp(x)
```

```
##
## Classification tree:
## rpart(formula = readmitted ~ ., data = train, control = rpart.control(cp = 0,
##     minsplit = 30, xval = 10))
##
## Variables actually used in tree construction:
##  [1] A1Cresult.7              A1Cresult.8              A1CresultNone
##  [4] A1CresultNorm            admission_source_id      admission_type_id
##  [7] age.30.40.               age.40.50.               age.50.60.
## [10] age.60.70.               age.70.80.               age.80.90.
## [13] age.90.100.              changeCh                 changeNo
## [16] diabetesMedNo            diabetesMedYes           diag_1
## [19] diag_2                   diag_3                   discharge_disposition_id
## [22] encounter_id             genderFemale             genderMale
## [25] glimepirideNo            glimepirideSteady        glipizideNo
## [28] glipizideSteady          glipizideUp              glyburideNo
## [31] glyburideSteady          insulinDown              insulinNo
## [34] insulinSteady            insulinUp                max_glu_serum.200
## [37] max_glu_serumNone        max_glu_serumNorm        metforminNo
## [40] metforminSteady          num_lab_procedures       num_medications
## [43] num_procedures           number_diagnoses         number_emergency
## [46] number_inpatient         number_outpatient        patient_nbr
## [49] pioglitazoneNo           pioglitazoneSteady       raceAfricanAmerican
## [52] raceCaucasian            raceHispanic             raceOther
## [55] rosiglitazoneNo          rosiglitazoneSteady      time_in_hospital
##
## Root node error: 32096/68637 = 0.46762
##
## n= 68637
##
##             CP nsplit rel error  xerror      xstd
## 1   1.6644e-01      0   1.00000 1.00000 0.0040727
## 2   7.8099e-03      1   0.83356 0.83356 0.0039809
## 3   4.6501e-03      4   0.81013 0.81303 0.0039624
## 4   3.9413e-03     10   0.78203 0.78742 0.0039370
## 5   2.7106e-03     12   0.77415 0.78374 0.0039331
## 6   2.6795e-03     13   0.77144 0.78150 0.0039307
## 7   2.3056e-03     14   0.76876 0.78035 0.0039295
## 8   2.2744e-03     15   0.76645 0.77854 0.0039276
## 9   1.9940e-03     16   0.76418 0.77736 0.0039263
## 10  1.9005e-03     17   0.76218 0.77639 0.0039252
## 11  1.6046e-03     18   0.76028 0.77265 0.0039211
## 12  1.4176e-03     20   0.75707 0.77119 0.0039195
## 13  1.3397e-03     22   0.75424 0.77103 0.0039193
## 14  1.1372e-03     23   0.75290 0.77022 0.0039185
## 15  9.8143e-04     25   0.75062 0.76679 0.0039146
## 16  6.6987e-04     28   0.74704 0.76315 0.0039105
## 17  6.3871e-04     30   0.74570 0.76022 0.0039071
## 18  5.9197e-04     32   0.74442 0.76044 0.0039074
## 19  5.8159e-04     33   0.74383 0.76006 0.0039069
## 20  5.5043e-04     37   0.74115 0.75969 0.0039065
## 21  5.1928e-04     40   0.73950 0.75997 0.0039068
## 22  4.6735e-04     43   0.73794 0.75963 0.0039064
## 23  4.5696e-04     44   0.73748 0.75928 0.0039060
## 24  4.3619e-04     51   0.73377 0.75804 0.0039046
## 25  4.0503e-04     52   0.73333 0.75785 0.0039044
## 26  3.9465e-04     56   0.73171 0.75735 0.0039038
## 27  3.7388e-04     59   0.73053 0.75732 0.0039038
```

```
## 28  3.4272e-04     64   0.72866 0.75704 0.0039034
## 29  3.3493e-04     66   0.72797 0.75654 0.0039029
## 30  3.2714e-04     72   0.72588 0.75651 0.0039028
## 31  3.1157e-04     76   0.72458 0.75679 0.0039032
## 32  2.9599e-04     97   0.71769 0.75695 0.0039033
## 33  2.9287e-04    104   0.71532 0.75664 0.0039030
## 34  2.9079e-04    113   0.71218 0.75629 0.0039026
## 35  2.8041e-04    116   0.71130 0.75645 0.0039028
## 36  2.7002e-04    125   0.70878 0.75751 0.0039040
## 37  2.6483e-04    130   0.70722 0.75763 0.0039041
## 38  2.5704e-04    135   0.70588 0.75720 0.0039036
## 39  2.4925e-04    143   0.70317 0.75732 0.0039038
## 40  2.3887e-04    157   0.69968 0.75729 0.0039037
## 41  2.2700e-04    171   0.69622 0.75732 0.0039038
## 42  2.2588e-04    183   0.69295 0.75760 0.0039041
## 43  2.1810e-04    189   0.69152 0.75810 0.0039047
## 44  2.0771e-04    214   0.68597 0.75770 0.0039042
## 45  2.0563e-04    221   0.68445 0.75726 0.0039037
## 46  2.0252e-04    228   0.68298 0.75726 0.0039037
## 47  1.9732e-04    240   0.68043 0.75726 0.0039037
## 48  1.9584e-04    247   0.67899 0.75726 0.0039037
## 49  1.9317e-04    260   0.67569 0.75689 0.0039033
## 50  1.8694e-04    268   0.67392 0.75832 0.0039049
## 51  1.7759e-04    302   0.66659 0.75872 0.0039054
## 52  1.7136e-04    317   0.66363 0.75950 0.0039063
## 53  1.6617e-04    341   0.65930 0.76203 0.0039092
## 54  1.6357e-04    373   0.65295 0.76221 0.0039094
## 55  1.6201e-04    384   0.65074 0.76284 0.0039101
## 56  1.5578e-04    390   0.64974 0.76402 0.0039115
## 57  1.4799e-04    471   0.63494 0.76577 0.0039135
## 58  1.4540e-04    479   0.63372 0.76642 0.0039142
## 59  1.4020e-04    485   0.63285 0.76673 0.0039145
## 60  1.3501e-04    526   0.62662 0.76932 0.0039174
## 61  1.3086e-04    537   0.62500 0.76932 0.0039174
## 62  1.2463e-04    546   0.62354 0.77109 0.0039194
## 63  1.1770e-04    628   0.61269 0.77299 0.0039215
## 64  1.1684e-04    641   0.61104 0.77346 0.0039220
## 65  1.1424e-04    654   0.60939 0.77346 0.0039220
## 66  1.1294e-04    680   0.60609 0.77483 0.0039235
## 67  1.0905e-04    701   0.60303 0.77536 0.0039241
## 68  1.0386e-04    759   0.59612 0.77586 0.0039247
## 69  9.8662e-05    811   0.58948 0.78060 0.0039298
## 70  9.3470e-05    832   0.58724 0.78137 0.0039306
## 71  8.7238e-05    986   0.57191 0.78499 0.0039344
## 72  8.5680e-05    996   0.57094 0.78876 0.0039384
## 73  8.4972e-05   1005   0.57013 0.78876 0.0039384
## 74  8.4568e-05   1025   0.56780 0.78876 0.0039384
## 75  8.3084e-05   1032   0.56720 0.78879 0.0039384
## 76  7.7891e-05   1091   0.56166 0.79446 0.0039442
## 77  7.4776e-05   1199   0.55241 0.79561 0.0039454
## 78  7.2699e-05   1217   0.55063 0.79677 0.0039466
## 79  7.0102e-05   1250   0.54792 0.79683 0.0039466
## 80  6.8544e-05   1277   0.54580 0.79683 0.0039466
## 81  6.6208e-05   1284   0.54527 0.80558 0.0039553
## 82  6.2313e-05   1309   0.54318 0.80630 0.0039560
## 83  5.7120e-05   1578   0.52427 0.80976 0.0039593
## 84  5.6082e-05   1584   0.52393 0.81540 0.0039646
## 85  5.4524e-05   1591   0.52331 0.81680 0.0039659
## 86  5.1928e-05   1613   0.52178 0.81708 0.0039662
```

```
## 87  4.9850e-05   1686   0.51633 0.82244 0.0039711
## 88  4.6735e-05   1702   0.51552 0.82412 0.0039726
## 89  4.4509e-05   1831   0.50791 0.82630 0.0039746
## 90  4.3619e-05   1856   0.50589 0.82633 0.0039746
## 91  4.1542e-05   1910   0.50215 0.82752 0.0039756
## 92  4.0058e-05   1955   0.50009 0.82908 0.0039770
## 93  3.8946e-05   1967   0.49944 0.83013 0.0039779
## 94  3.7388e-05   1989   0.49857 0.83020 0.0039780
## 95  3.1157e-05   1999   0.49819 0.84169 0.0039878
## 96  2.4925e-05   2184   0.49143 0.84244 0.0039884
## 97  2.3367e-05   2190   0.49128 0.84425 0.0039899
## 98  2.0771e-05   2198   0.49109 0.84652 0.0039918
## 99  1.8694e-05   2252   0.48994 0.84746 0.0039925
## 100 1.7804e-05   2267   0.48966 0.84746 0.0039925
## 101 1.5578e-05   2274   0.48953 0.85051 0.0039950
## 102 1.0386e-05   2343   0.48841 0.85123 0.0039956
## 103 7.7891e-06   2358   0.48825 0.85360 0.0039974
## 104 6.2313e-06   2370   0.48816 0.85394 0.0039977
## 105 0.0000e+00   2375   0.48813 0.85459 0.0039982
```

```
# 2.3. Plot complexity parameter
plotcp (x)
```




```
# 2.4. Find lowest xerror value and build a pruned tree using that cp value as an argument
in the r part formula.
# Find lowest xerror value and cp
min_xerror <- which.min(x$cptable[,"xerror"])
cp <- x$cptable[min_xerror, "CP"]
print(cp)
```

```
## [1] 0.0002907943
```

```
#build a pruned tree using that cp value as an argument
x_pruned=rpart(readmitted~., data=train,control=rpart.control(cp=0.0003349327,minsplit=30,
xval=10))

printcp(x_pruned)
```

```
##
## Classification tree:
## rpart(formula = readmitted ~ ., data = train, control = rpart.control(cp = 0.000334932
7,
##     minsplit = 30, xval = 10))
##
## Variables actually used in tree construction:
##  [1] A1CresultNone            admission_source_id      admission_type_id
##  [4] age.70.80.               age.90.100.              diabetesMedNo
##  [7] diag_1                   diag_3                   discharge_disposition_id
## [10] encounter_id             num_lab_procedures       num_medications
## [13] num_procedures           number_diagnoses         number_emergency
## [16] number_inpatient         number_outpatient        patient_nbr
## [19] raceAfricanAmerican      time_in_hospital
##
## Root node error: 32096/68637 = 0.46762
##
## n= 68637
##
##            CP nsplit rel error  xerror      xstd
## 1  0.16643819      0   1.00000 1.00000 0.0040727
## 2  0.00780990      1   0.83356 0.83356 0.0039809
## 3  0.00465011      4   0.81013 0.81119 0.0039607
## 4  0.00394130     10   0.78203 0.78218 0.0039315
## 5  0.00271062     12   0.77415 0.78032 0.0039295
## 6  0.00267946     13   0.77144 0.78159 0.0039308
## 7  0.00230558     14   0.76876 0.77991 0.0039290
## 8  0.00227443     15   0.76645 0.77954 0.0039286
## 9  0.00199402     16   0.76418 0.77810 0.0039271
## 10 0.00190055     17   0.76218 0.77499 0.0039237
## 11 0.00160456     18   0.76028 0.77069 0.0039190
## 12 0.00141762     20   0.75707 0.76991 0.0039181
## 13 0.00133973     22   0.75424 0.76919 0.0039173
## 14 0.00113721     23   0.75290 0.76658 0.0039144
## 15 0.00098143     25   0.75062 0.76390 0.0039113
## 16 0.00066987     28   0.74704 0.76059 0.0039076
## 17 0.00063871     30   0.74570 0.75891 0.0039056
## 18 0.00059197     32   0.74442 0.75841 0.0039050
## 19 0.00058159     33   0.74383 0.75707 0.0039035
## 20 0.00055043     37   0.74115 0.75757 0.0039041
## 21 0.00051928     40   0.73950 0.75598 0.0039022
## 22 0.00046735     43   0.73794 0.75580 0.0039020
## 23 0.00045696     44   0.73748 0.75411 0.0039000
## 24 0.00043619     51   0.73377 0.75480 0.0039008
## 25 0.00040503     52   0.73333 0.75523 0.0039013
## 26 0.00039465     56   0.73171 0.75467 0.0039007
## 27 0.00037388     59   0.73053 0.75486 0.0039009
## 28 0.00034272     64   0.72866 0.75536 0.0039015
## 29 0.00033493     66   0.72797 0.75561 0.0039018
## 30 0.00033493     72   0.72588 0.75561 0.0039018
```

```
# 2.5. Generate confusion matrix for pruned tree
Prediction_Model<-predict(x_pruned, data=train, type="class")
confusionMatrix(Prediction_Model, train$readmitted)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction     1     0
##          1 19329 10531
##          0 12767 26010
##
##                Accuracy : 0.6606
##                  95% CI : (0.657, 0.6641)
##     No Information Rate : 0.5324
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.3154
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.6022
##             Specificity : 0.7118
##          Pos Pred Value : 0.6473
##          Neg Pred Value : 0.6708
##              Prevalence : 0.4676
##          Detection Rate : 0.2816
##    Detection Prevalence : 0.4350
##       Balanced Accuracy : 0.6570
##
##        'Positive' Class : 1
##
```

```
round(prop.table(table(Prediction_Model,train$readmitted),1),2)
```

```
##
## Prediction_Model    1    0
##                1 0.65 0.35
##                0 0.33 0.67
```

```
#sapply(Prediction_Model, levels)

# 2.6. Point out significant interactions you think you see
#The numbers of TP and TN are almost the same, the numbers of FP and FN are almost half of numbers of TP and TN
summary(x_pruned)
```

```
## Call:
## rpart(formula = readmitted ~ ., data = train, control = rpart.control(cp = 0.000334932
7,
##     minsplit = 30, xval = 10))
##   n= 68637
##
##              CP nsplit rel error    xerror        xstd
## 1  0.1664381854      0 1.0000000 1.0000000 0.004072730
## 2  0.0078099036      1 0.8335618 0.8335618 0.003980918
## 3  0.0046501122      4 0.8101321 0.8111914 0.003960655
## 4  0.0039413011     10 0.7820289 0.7821847 0.003931465
## 5  0.0027106181     12 0.7741463 0.7803153 0.003929469
## 6  0.0026794616     13 0.7714357 0.7815927 0.003930834
## 7  0.0023055833     14 0.7687562 0.7799103 0.003929035
## 8  0.0022744267     15 0.7664506 0.7795364 0.003928633
## 9  0.0019940179     16 0.7641762 0.7781032 0.003927089
## 10 0.0019005484     17 0.7621822 0.7749875 0.003923705
## 11 0.0016045613     18 0.7602817 0.7706879 0.003918970
## 12 0.0014176221     20 0.7570725 0.7699090 0.003918104
## 13 0.0013397308     22 0.7542373 0.7691924 0.003917305
## 14 0.0011372134     23 0.7528976 0.7665753 0.003914371
## 15 0.0009814307     25 0.7506231 0.7638958 0.003911338
## 16 0.0006698654     28 0.7470401 0.7605932 0.003907559
## 17 0.0006387089     30 0.7457004 0.7589108 0.003905617
## 18 0.0005919741     32 0.7444230 0.7584123 0.003905039
## 19 0.0005815886     33 0.7438310 0.7570725 0.003903482
## 20 0.0005504320     37 0.7411515 0.7575710 0.003904062
## 21 0.0005192755     40 0.7395002 0.7559821 0.003902209
## 22 0.0004673480     43 0.7379424 0.7557951 0.003901990
## 23 0.0004569624     44 0.7374751 0.7541127 0.003900015
## 24 0.0004361914     51 0.7337674 0.7547981 0.003900821
## 25 0.0004050349     52 0.7333313 0.7552343 0.003901333
## 26 0.0003946494     56 0.7317111 0.7546735 0.003900675
## 27 0.0003738784     59 0.7305272 0.7548604 0.003900895
## 28 0.0003427218     64 0.7286578 0.7553589 0.003901479
## 29 0.0003349327     66 0.7279723 0.7556082 0.003901772
## 30 0.0003349327     72 0.7258848 0.7556082 0.003901772
##
## Variable importance
##         number_inpatient             encounter_id              patient_nbr
##                       35                       18                       13
## discharge_disposition_id         number_emergency         number_diagnoses
##                       12                        4                        4
##      admission_source_id                   diag_1       num_lab_procedures
##                        2                        1                        1
##        number_outpatient            diabetesMedNo           diabetesMedYes
##                        1                        1                        1
##          num_medications                   diag_3         time_in_hospital
##                        1                        1                        1
##        admission_type_id        max_glu_serumNone
##                        1                        1
##
## Node number 1: 68637 observations,    complexity param=0.1664382
##   predicted class=0  expected loss=0.4676195  P(node) =1
##     class counts: 32096 36541
##    probabilities: 0.468 0.532
##   left son=2 (23320 obs) right son=3 (45317 obs)
##   Primary splits:
```

```
##       number_inpatient  < 0.5       to the right, improve=1524.7620, (0 missing)
##       number_emergency  < 0.5       to the right, improve= 493.4800, (0 missing)
##       patient_nbr       < 35598660  to the right, improve= 433.2063, (0 missing)
##       number_outpatient < 0.5       to the right, improve= 390.1303, (0 missing)
##       number_diagnoses  < 6.5       to the right, improve= 332.8270, (0 missing)
##   Surrogate splits:
##       number_emergency    < 0.5       to the right, agree=0.683, adj=0.067, (0 split)
##       diag_1              < 995.5     to the right, agree=0.662, adj=0.007, (0 split)
##       number_outpatient   < 4.5       to the right, agree=0.662, adj=0.006, (0 split)
##       diag_2              < 8.5       to the left,  agree=0.660, adj=0.000, (0 split)
##       admission_source_id < 21        to the right, agree=0.660, adj=0.000, (0 split)
##
## Node number 2: 23320 observations,    complexity param=0.004650112
##   predicted class=1  expected loss=0.3854631  P(node) =0.3397584
##     class counts: 14331  8989
##    probabilities: 0.615 0.385
##   left son=4 (10014 obs) right son=5 (13306 obs)
##   Primary splits:
##       number_inpatient         < 1.5       to the right, improve=250.53740, (0 missing)
##       discharge_disposition_id < 9.5       to the left,  improve=176.92940, (0 missing)
##       number_emergency         < 0.5       to the right, improve= 97.47924, (0 missing)
##       patient_nbr              < 29208760  to the right, improve= 91.73974, (0 missing)
##       number_outpatient        < 0.5       to the right, improve= 77.59470, (0 missing)
##   Surrogate splits:
##       number_emergency  < 0.5       to the right, agree=0.599, adj=0.065, (0 split)
##       number_outpatient < 0.5       to the right, agree=0.577, adj=0.015, (0 split)
##       age.20.30.        < 0.5       to the right, agree=0.574, adj=0.008, (0 split)
##       age.30.40.        < 0.5       to the right, agree=0.572, adj=0.004, (0 split)
##       patient_nbr       < 107734.5  to the left,  agree=0.571, adj=0.001, (0 split)
##
## Node number 3: 45317 observations,    complexity param=0.007809904
##   predicted class=0  expected loss=0.3920162  P(node) =0.6602416
##     class counts: 17765 27552
##    probabilities: 0.392 0.608
##   left son=6 (26187 obs) right son=7 (19130 obs)
##   Primary splits:
##       patient_nbr       < 40364820  to the right, improve=272.9472, (0 missing)
##       number_diagnoses  < 6.5       to the right, improve=198.3098, (0 missing)
##       encounter_id      < 286017200 to the left,  improve=150.0056, (0 missing)
##       number_outpatient < 0.5       to the right, improve=130.9355, (0 missing)
##       number_emergency  < 0.5       to the right, improve=125.9344, (0 missing)
##   Surrogate splits:
##       encounter_id             < 120386500 to the right, agree=0.792, adj=0.507, (0 spl
it)
##       number_diagnoses         < 5.5       to the right, agree=0.623, adj=0.107, (0 spl
it)
##       discharge_disposition_id < 15.5      to the left,  agree=0.615, adj=0.087, (0 spl
it)
##       max_glu_serumNone        < 0.5       to the right, agree=0.610, adj=0.075, (0 spl
it)
##       admission_source_id      < 7.5       to the left,  agree=0.609, adj=0.073, (0 spl
it)
##
## Node number 4: 10014 observations,    complexity param=0.004650112
##   predicted class=1  expected loss=0.3009786  P(node) =0.145898
##     class counts:  7000  3014
##    probabilities: 0.699 0.301
##   left son=8 (8901 obs) right son=9 (1113 obs)
##   Primary splits:
```

```
##       discharge_disposition_id < 10.5      to the left,  improve=123.34890, (0 missing)
##       number_inpatient         < 3.5       to the right, improve= 69.71732, (0 missing)
##       patient_nbr              < 29891380  to the right, improve= 37.80872, (0 missing)
##       number_emergency         < 1.5       to the right, improve= 36.82188, (0 missing)
##       number_outpatient        < 0.5       to the right, improve= 25.61829, (0 missing)
##   Surrogate splits:
##       tolazamideNo     < 0.5       to the right, agree=0.889, adj=0.003, (0 split)
##       tolazamideSteady < 0.5       to the left,  agree=0.889, adj=0.003, (0 split)
##       num_medications  < 53.5      to the left,  agree=0.889, adj=0.002, (0 split)
##
## Node number 5: 13306 observations,    complexity param=0.004650112
##   predicted class=1  expected loss=0.4490455  P(node) =0.1938605
##     class counts:  7331  5975
##    probabilities: 0.551 0.449
##   left son=10 (12002 obs) right son=11 (1304 obs)
##   Primary splits:
##       discharge_disposition_id < 7.5       to the left,  improve=70.37829, (0 missing)
##       patient_nbr              < 26551820  to the right, improve=50.92464, (0 missing)
##       encounter_id             < 271173400 to the left,  improve=46.89946, (0 missing)
##       admission_source_id      < 6.5       to the right, improve=38.97114, (0 missing)
##       number_outpatient        < 0.5       to the right, improve=28.24145, (0 missing)
##   Surrogate splits:
##       encounter_id < 7402494   to the right, agree=0.902, adj=0.001, (0 split)
##
## Node number 6: 26187 observations,    complexity param=0.007809904
##   predicted class=0  expected loss=0.4389201  P(node) =0.3815289
##     class counts: 11494 14693
##    probabilities: 0.439 0.561
##   left son=12 (15803 obs) right son=13 (10384 obs)
##   Primary splits:
##       encounter_id     < 215571600 to the left,  improve=376.93600, (0 missing)
##       patient_nbr      < 115183600 to the left,  improve=189.46180, (0 missing)
##       number_emergency < 0.5       to the right, improve= 71.46888, (0 missing)
##       diabetesMedNo    < 0.5       to the left,  improve= 69.96509, (0 missing)
##       diabetesMedYes   < 0.5       to the right, improve= 69.96509, (0 missing)
##   Surrogate splits:
##       patient_nbr        < 115176100 to the left,  agree=0.676, adj=0.182, (0 split)
##       num_lab_procedures < 1.5       to the right, agree=0.606, adj=0.007, (0 split)
##       number_diagnoses   < 9.5       to the left,  agree=0.605, adj=0.003, (0 split)
##       raceAsian          < 0.5       to the left,  agree=0.604, adj=0.002, (0 split)
##       num_procedures     < 5.5       to the left,  agree=0.604, adj=0.002, (0 split)
##
## Node number 7: 19130 observations,    complexity param=0.0004569624
##   predicted class=0  expected loss=0.3278097  P(node) =0.2787126
##     class counts:  6271 12859
##    probabilities: 0.328 0.672
##   left son=14 (10930 obs) right son=15 (8200 obs)
##   Primary splits:
##       number_diagnoses    < 6.5       to the right, improve=109.74760, (0 missing)
##       encounter_id        < 48352510  to the left,  improve= 82.50129, (0 missing)
##       admission_source_id < 6.5       to the right, improve= 71.57876, (0 missing)
##       number_emergency    < 0.5       to the right, improve= 37.77664, (0 missing)
##       diag_3              < 402.5     to the right, improve= 31.71590, (0 missing)
##   Surrogate splits:
##       diag_3              < 250.015   to the right, agree=0.634, adj=0.147, (0 split)
##       num_medications     < 10.5      to the right, agree=0.624, adj=0.124, (0 split)
##       diag_2              < 250.115   to the right, agree=0.617, adj=0.107, (0 split)
##       time_in_hospital    < 2.5       to the right, agree=0.604, adj=0.075, (0 split)
##       admission_source_id < 6.5       to the right, agree=0.601, adj=0.068, (0 split)
```

```
##
## Node number 8: 8901 observations
##   predicted class=1  expected loss=0.2732277  P(node) =0.1296822
##     class counts:  6469  2432
##    probabilities: 0.727 0.273
##
## Node number 9: 1113 observations,    complexity param=0.004650112
##   predicted class=0  expected loss=0.4770889  P(node) =0.01621574
##     class counts:   531   582
##    probabilities: 0.477 0.523
##   left son=18 (746 obs) right son=19 (367 obs)
##   Primary splits:
##       discharge_disposition_id < 14.5      to the right, improve=190.55630, (0 missing)
##       encounter_id             < 81935950  to the left,  improve= 36.39522, (0 missing)
##       patient_nbr              < 17811450  to the left,  improve= 34.57004, (0 missing)
##       diag_1                   < 55.5      to the right, improve= 20.66070, (0 missing)
##       num_lab_procedures       < 64.5      to the left,  improve= 16.52660, (0 missing)
##   Surrogate splits:
##       patient_nbr        < 44940020  to the left,  agree=0.734, adj=0.193, (0 split)
##       encounter_id       < 81596240  to the left,  agree=0.702, adj=0.095, (0 split)
##       diag_1             < 48.5      to the right, agree=0.701, adj=0.093, (0 split)
##       num_lab_procedures < 64.5      to the left,  agree=0.698, adj=0.084, (0 split)
##       time_in_hospital   < 1.5       to the right, agree=0.692, adj=0.065, (0 split)
##
## Node number 10: 12002 observations,    complexity param=0.004650112
##   predicted class=1  expected loss=0.4320947  P(node) =0.174862
##     class counts:  6816  5186
##    probabilities: 0.568 0.432
##   left son=20 (8728 obs) right son=21 (3274 obs)
##   Primary splits:
##       patient_nbr         < 26551820  to the right, improve=48.91977, (0 missing)
##       admission_source_id < 6.5       to the right, improve=44.65534, (0 missing)
##       encounter_id        < 271173400 to the left,  improve=44.64595, (0 missing)
##       number_emergency    < 0.5       to the right, improve=27.83131, (0 missing)
##       number_outpatient   < 0.5       to the right, improve=25.64781, (0 missing)
##   Surrogate splits:
##       encounter_id        < 117964000 to the right, agree=0.826, adj=0.363, (0 split)
##       max_glu_serumNone   < 0.5       to the right, agree=0.741, adj=0.051, (0 split)
##       max_glu_serumNorm   < 0.5       to the left,  agree=0.738, adj=0.041, (0 split)
##       admission_source_id < 8.5       to the left,  agree=0.737, adj=0.035, (0 split)
##       max_glu_serum.200   < 0.5       to the left,  agree=0.731, adj=0.013, (0 split)
##
## Node number 11: 1304 observations,    complexity param=0.002305583
##   predicted class=0  expected loss=0.3949387  P(node) =0.0189985
##     class counts:   515   789
##    probabilities: 0.395 0.605
##   left son=22 (916 obs) right son=23 (388 obs)
##   Primary splits:
##       discharge_disposition_id < 14.5      to the right, improve=130.26410, (0 missing)
##       patient_nbr              < 15436360  to the left,  improve= 28.39765, (0 missing)
##       encounter_id             < 82224980  to the left,  improve= 21.87664, (0 missing)
##       raceAfricanAmerican      < 0.5       to the right, improve= 12.05051, (0 missing)
##       raceCaucasian            < 0.5       to the left,  improve= 10.26051, (0 missing)
##   Surrogate splits:
##       patient_nbr        < 44039840  to the left,  agree=0.736, adj=0.113, (0 split)
##       encounter_id       < 219835500 to the left,  agree=0.729, adj=0.090, (0 split)
##       num_lab_procedures < 67.5      to the left,  agree=0.727, adj=0.082, (0 split)
##       diag_1             < 53.5      to the right, agree=0.712, adj=0.034, (0 split)
##       time_in_hospital   < 1.5       to the right, agree=0.706, adj=0.013, (0 split)
```

```
##
## Node number 12: 15803 observations,    complexity param=0.007809904
##   predicted class=1  expected loss=0.4923116  P(node) =0.2302402
##     class counts:  8023  7780
##    probabilities: 0.508 0.492
##   left son=24 (11908 obs) right son=25 (3895 obs)
##   Primary splits:
##       diabetesMedNo            < 0.5       to the left,  improve=55.13458, (0 missing)
##       diabetesMedYes           < 0.5       to the right, improve=55.13458, (0 missing)
##       discharge_disposition_id < 10.5      to the left,  improve=44.09366, (0 missing)
##       encounter_id             < 150861900 to the left,  improve=42.83677, (0 missing)
##       number_emergency         < 0.5       to the right, improve=39.76804, (0 missing)
##   Surrogate splits:
##       diabetesMedYes    < 0.5       to the right, agree=1.000, adj=1.000, (0 split)
##       num_medications   < 4.5       to the right, agree=0.763, adj=0.036, (0 split)
##       max_glu_serumNorm < 0.5       to the left,  agree=0.755, adj=0.005, (0 split)
##       diag_1            < 4         to the right, agree=0.754, adj=0.001, (0 split)
##
## Node number 13: 10384 observations,    complexity param=0.000550432
##   predicted class=0  expected loss=0.3342643  P(node) =0.1512887
##     class counts:  3471  6913
##    probabilities: 0.334 0.666
##   left son=26 (8495 obs) right son=27 (1889 obs)
##   Primary splits:
##       patient_nbr       < 115185200 to the left,  improve=56.76192, (0 missing)
##       number_emergency  < 0.5       to the right, improve=49.48911, (0 missing)
##       encounter_id      < 288191100 to the left,  improve=45.11556, (0 missing)
##       number_outpatient < 0.5       to the right, improve=30.30371, (0 missing)
##       diabetesMedNo     < 0.5       to the left,  improve=28.54275, (0 missing)
##   Surrogate splits:
##       raceOther    < 0.5       to the left,  agree=0.821, adj=0.018, (0 split)
##       encounter_id < 443712500 to the left,  agree=0.819, adj=0.004, (0 split)
##
## Node number 14: 10930 observations,    complexity param=0.0004569624
##   predicted class=0  expected loss=0.3741995  P(node) =0.1592436
##     class counts:  4090  6840
##    probabilities: 0.374 0.626
##   left son=28 (3162 obs) right son=29 (7768 obs)
##   Primary splits:
##       encounter_id        < 58293260  to the left,  improve=55.52634, (0 missing)
##       admission_source_id < 6.5       to the right, improve=16.12132, (0 missing)
##       number_emergency    < 0.5       to the right, improve=14.81437, (0 missing)
##       diabetesMedNo       < 0.5       to the left,  improve=13.28404, (0 missing)
##       diabetesMedYes      < 0.5       to the right, improve=13.28404, (0 missing)
##   Surrogate splits:
##       patient_nbr              < 19842120  to the left,  agree=0.761, adj=0.174, (0 spl
it)
##       discharge_disposition_id < 24.5      to the right, agree=0.719, adj=0.027, (0 spl
it)
##       admission_source_id      < 18.5      to the right, agree=0.714, adj=0.012, (0 spl
it)
##       acarboseUp               < 0.5       to the right, agree=0.711, adj=0.001, (0 spl
it)
##       troglitazoneNo           < 0.5       to the left,  agree=0.711, adj=0.001, (0 spl
it)
##
## Node number 15: 8200 observations
##   predicted class=0  expected loss=0.2659756  P(node) =0.1194691
##     class counts:  2181  6019
```

```
##    probabilities: 0.266 0.734
##
## Node number 18: 746 observations
##   predicted class=1  expected loss=0.3176944  P(node) =0.01086877
##     class counts:   509   237
##    probabilities: 0.682 0.318
##
## Node number 19: 367 observations
##   predicted class=0  expected loss=0.0599455  P(node) =0.00534697
##     class counts:    22   345
##    probabilities: 0.060 0.940
##
## Node number 20: 8728 observations,    complexity param=0.002679462
##   predicted class=1  expected loss=0.4044455  P(node) =0.1271617
##     class counts:  5198  3530
##    probabilities: 0.596 0.404
##   left son=40 (6448 obs) right son=41 (2280 obs)
##   Primary splits:
##       encounter_id        < 261755100 to the left,  improve=80.80055, (0 missing)
##       admission_source_id < 6.5       to the right, improve=23.39647, (0 missing)
##       patient_nbr         < 123699400 to the left,  improve=18.20816, (0 missing)
##       number_emergency    < 0.5       to the right, improve=15.89227, (0 missing)
##       num_lab_procedures  < 35.5      to the right, improve=13.95577, (0 missing)
##   Surrogate splits:
##       patient_nbr         < 117688300 to the left,  agree=0.763, adj=0.094, (0 split)
##       number_diagnoses    < 9.5       to the left,  agree=0.741, adj=0.008, (0 split)
##       admission_source_id < 21        to the left,  agree=0.739, adj=0.003, (0 split)
##       num_medications     < 51.5      to the left,  agree=0.739, adj=0.001, (0 split)
##       nateglinideDown     < 0.5       to the left,  agree=0.739, adj=0.001, (0 split)
##
## Node number 21: 3274 observations,    complexity param=0.004650112
##   predicted class=0  expected loss=0.4941967  P(node) =0.04770022
##     class counts:  1618  1656
##    probabilities: 0.494 0.506
##   left son=42 (1201 obs) right son=43 (2073 obs)
##   Primary splits:
##       encounter_id        < 68341090  to the left,  improve=51.895780, (0 missing)
##       admission_source_id < 6.5       to the right, improve=18.344640, (0 missing)
##       number_diagnoses    < 5.5       to the right, improve=16.485930, (0 missing)
##       patient_nbr         < 2401816   to the left,  improve= 9.195141, (0 missing)
##       number_emergency    < 0.5       to the right, improve= 8.647468, (0 missing)
##   Surrogate splits:
##       patient_nbr         < 5176755   to the left,  agree=0.695, adj=0.167, (0 split)
##       num_lab_procedures  < 53.5      to the right, agree=0.644, adj=0.030, (0 split)
##       admission_source_id < 18.5      to the right, agree=0.636, adj=0.008, (0 split)
##       diag_2              < 884.5     to the right, agree=0.635, adj=0.005, (0 split)
##       raceOther           < 0.5       to the right, agree=0.635, adj=0.004, (0 split)
##
## Node number 22: 916 observations,    complexity param=0.0006698654
##   predicted class=1  expected loss=0.459607  P(node) =0.01334557
##     class counts:   495   421
##    probabilities: 0.540 0.460
##   left son=44 (250 obs) right son=45 (666 obs)
##   Primary splits:
##       raceAfricanAmerican < 0.5       to the right, improve=6.822929, (0 missing)
##       raceCaucasian       < 0.5       to the left,  improve=4.730564, (0 missing)
##       patient_nbr         < 3788996   to the left,  improve=4.574157, (0 missing)
##       time_in_hospital    < 5.5       to the left,  improve=3.583783, (0 missing)
##       diag_3              < 995.5     to the right, improve=2.614891, (0 missing)
```

```
##   Surrogate splits:
##       raceCaucasian   < 0.5       to the left,  agree=0.963, adj=0.864, (0 split)
##       patient_nbr     < 220909.5  to the left,  agree=0.735, adj=0.028, (0 split)
##       glimepirideDown < 0.5       to the right, agree=0.728, adj=0.004, (0 split)
##
## Node number 23: 388 observations
##   predicted class=0  expected loss=0.05154639  P(node) =0.005652928
##     class counts:    20   368
##    probabilities: 0.052 0.948
##
## Node number 24: 11908 observations,    complexity param=0.003941301
##   predicted class=1  expected loss=0.4684246  P(node) =0.1734924
##     class counts:  6330  5578
##    probabilities: 0.532 0.468
##   left son=48 (6169 obs) right son=49 (5739 obs)
##   Primary splits:
##       encounter_id             < 153802300 to the left,  improve=31.88451, (0 missing)
##       discharge_disposition_id < 6.5       to the left,  improve=28.86864, (0 missing)
##       number_outpatient        < 0.5       to the right, improve=23.75389, (0 missing)
##       number_emergency         < 0.5       to the right, improve=21.30748, (0 missing)
##       number_diagnoses         < 6.5       to the right, improve=20.35582, (0 missing)
##   Surrogate splits:
##       number_diagnoses  < 8.5       to the left,  agree=0.566, adj=0.099, (0 split)
##       patient_nbr       < 60535820  to the right, agree=0.558, adj=0.083, (0 split)
##       A1CresultNone     < 0.5       to the right, agree=0.554, adj=0.074, (0 split)
##       num_medications   < 17.5      to the left,  agree=0.551, adj=0.068, (0 split)
##       number_outpatient < 0.5       to the left,  agree=0.544, adj=0.053, (0 split)
##
## Node number 25: 3895 observations,    complexity param=0.001900548
##   predicted class=0  expected loss=0.4346598  P(node) =0.05674782
##     class counts:  1693  2202
##    probabilities: 0.435 0.565
##   left son=50 (223 obs) right son=51 (3672 obs)
##   Primary splits:
##       number_emergency         < 0.5       to the right, improve=19.32509, (0 missing)
##       encounter_id             < 151601100 to the left,  improve=16.56320, (0 missing)
##       discharge_disposition_id < 9.5       to the left,  improve=15.14567, (0 missing)
##       number_outpatient        < 0.5       to the right, improve=11.31313, (0 missing)
##       number_diagnoses         < 7.5       to the right, improve=11.08976, (0 missing)
##
## Node number 26: 8495 observations,    complexity param=0.000550432
##   predicted class=0  expected loss=0.358917  P(node) =0.1237671
##     class counts:  3049  5446
##    probabilities: 0.359 0.641
##   left son=52 (911 obs) right son=53 (7584 obs)
##   Primary splits:
##       number_emergency < 0.5       to the right, improve=40.30669, (0 missing)
##       diabetesMedYes   < 0.5       to the right, improve=30.45079, (0 missing)
##       diabetesMedNo    < 0.5       to the left,  improve=30.45079, (0 missing)
##       patient_nbr      < 83703860  to the right, improve=20.08703, (0 missing)
##       num_medications  < 7.5       to the right, improve=20.05686, (0 missing)
##   Surrogate splits:
##       num_lab_procedures < 113.5     to the right, agree=0.893, adj=0.002, (0 split)
##
## Node number 27: 1889 observations
##   predicted class=0  expected loss=0.2233986  P(node) =0.0275216
##     class counts:   422  1467
##    probabilities: 0.223 0.777
##
```

```
## Node number 28: 3162 observations,    complexity param=0.0004569624
##   predicted class=0  expected loss=0.4531942  P(node) =0.04606845
##     class counts:  1433  1729
##    probabilities: 0.453 0.547
##   left son=56 (2071 obs) right son=57 (1091 obs)
##   Primary splits:
##       admission_source_id      < 6.5       to the right, improve=19.484170, (0 missing)
##       number_outpatient        < 0.5       to the right, improve= 8.542642, (0 missing)
##       patient_nbr              < 3982455   to the left,  improve= 6.786369, (0 missing)
##       encounter_id             < 30571210  to the left,  improve= 6.500032, (0 missing)
##       discharge_disposition_id < 6.5       to the left,  improve= 6.047995, (0 missing)
##   Surrogate splits:
##       admission_type_id  < 1.5       to the left,  agree=0.786, adj=0.379, (0 split)
##       num_procedures     < 4.5       to the left,  agree=0.680, adj=0.073, (0 split)
##       num_medications    < 26.5      to the left,  agree=0.673, adj=0.052, (0 split)
##       num_lab_procedures < 15.5      to the right, agree=0.665, adj=0.030, (0 split)
##       diag_1             < 236.5     to the right, agree=0.661, adj=0.017, (0 split)
##
## Node number 29: 7768 observations
##   predicted class=0  expected loss=0.3420443  P(node) =0.1131751
##     class counts:  2657  5111
##    probabilities: 0.342 0.658
##
## Node number 40: 6448 observations,    complexity param=0.0003946494
##   predicted class=1  expected loss=0.3639888  P(node) =0.0939435
##     class counts:  4101  2347
##    probabilities: 0.636 0.364
##   left son=80 (4221 obs) right son=81 (2227 obs)
##   Primary splits:
##       admission_source_id < 6.5       to the right, improve=16.119550, (0 missing)
##       number_emergency    < 1.5       to the right, improve=12.074590, (0 missing)
##       number_outpatient   < 0.5       to the right, improve=10.572550, (0 missing)
##       encounter_id        < 177251600 to the left,  improve= 9.955730, (0 missing)
##       diag_1              < 250.015   to the right, improve= 9.906099, (0 missing)
##   Surrogate splits:
##       admission_type_id  < 1.5       to the left,  agree=0.780, adj=0.362, (0 split)
##       diag_1             < 247.5     to the right, agree=0.686, adj=0.090, (0 split)
##       num_lab_procedures < 34.5      to the right, agree=0.664, adj=0.026, (0 split)
##       time_in_hospital   < 13.5      to the left,  agree=0.657, adj=0.006, (0 split)
##       num_medications    < 34.5      to the left,  agree=0.656, adj=0.004, (0 split)
##
## Node number 41: 2280 observations,    complexity param=0.001417622
##   predicted class=0  expected loss=0.4811404  P(node) =0.03321824
##     class counts:  1097  1183
##    probabilities: 0.481 0.519
##   left son=82 (1755 obs) right son=83 (525 obs)
##   Primary splits:
##       admission_type_id   < 2.5       to the left,  improve=14.217940, (0 missing)
##       admission_source_id < 6.5       to the right, improve=11.563240, (0 missing)
##       number_emergency    < 0.5       to the right, improve=10.773360, (0 missing)
##       num_lab_procedures  < 33.5      to the right, improve= 8.942285, (0 missing)
##       num_medications     < 10.5      to the right, improve= 8.157404, (0 missing)
##   Surrogate splits:
##       admission_source_id < 1.5       to the right, agree=0.898, adj=0.556, (0 split)
##       num_medications     < 36.5      to the left,  agree=0.777, adj=0.030, (0 split)
##       max_glu_serumNorm   < 0.5       to the left,  agree=0.771, adj=0.008, (0 split)
##       diag_3              < 995.5     to the left,  agree=0.771, adj=0.008, (0 split)
##       max_glu_serum.200   < 0.5       to the left,  agree=0.771, adj=0.006, (0 split)
##
```

```
## Node number 42: 1201 observations
##   predicted class=1  expected loss=0.3888426  P(node) =0.01749785
##     class counts:   734   467
##    probabilities: 0.611 0.389
##
## Node number 43: 2073 observations,    complexity param=0.0009814307
##   predicted class=0  expected loss=0.4264351  P(node) =0.03020237
##     class counts:   884  1189
##    probabilities: 0.426 0.574
##   left son=86 (1091 obs) right son=87 (982 obs)
##   Primary splits:
##       patient_nbr         < 21595690  to the right, improve=19.927340, (0 missing)
##       number_emergency    < 0.5       to the right, improve=17.198080, (0 missing)
##       admission_source_id < 6.5       to the right, improve=12.461370, (0 missing)
##       raceAfricanAmerican < 0.5       to the right, improve= 9.917530, (0 missing)
##       raceCaucasian       < 0.5       to the left,  improve= 9.874403, (0 missing)
##   Surrogate splits:
##       encounter_id        < 101461000 to the right, agree=0.767, adj=0.508, (0 split)
##       num_lab_procedures  < 46.5      to the left,  agree=0.674, adj=0.312, (0 split)
##       admission_source_id < 6.5       to the right, agree=0.633, adj=0.225, (0 split)
##       number_diagnoses    < 5.5       to the right, agree=0.618, adj=0.195, (0 split)
##       changeCh            < 0.5       to the right, agree=0.607, adj=0.170, (0 split)
##
## Node number 44: 250 observations
##   predicted class=1  expected loss=0.36  P(node) =0.00364235
##     class counts:   160    90
##    probabilities: 0.640 0.360
##
## Node number 45: 666 observations,    complexity param=0.0006698654
##   predicted class=1  expected loss=0.496997  P(node) =0.009703221
##     class counts:   335   331
##    probabilities: 0.503 0.497
##   left son=90 (293 obs) right son=91 (373 obs)
##   Primary splits:
##       encounter_id             < 71727880  to the right, improve=6.236165, (0 missing)
##       patient_nbr              < 26937110  to the right, improve=5.104141, (0 missing)
##       A1CresultNorm            < 0.5       to the right, improve=2.707189, (0 missing)
##       discharge_disposition_id < 16.5      to the left,  improve=2.666887, (0 missing)
##       number_outpatient        < 0.5       to the right, improve=2.646190, (0 missing)
##   Surrogate splits:
##       patient_nbr              < 20630890  to the right, agree=0.974, adj=0.942, (0 spl
it)
##       discharge_disposition_id < 20        to the right, agree=0.922, adj=0.823, (0 spl
it)
##       insulinNo                < 0.5       to the left,  agree=0.701, adj=0.321, (0 spl
it)
##       changeCh                 < 0.5       to the right, agree=0.661, adj=0.229, (0 spl
it)
##       changeNo                 < 0.5       to the left,  agree=0.661, adj=0.229, (0 spl
it)
##
## Node number 48: 6169 observations,    complexity param=0.001994018
##   predicted class=1  expected loss=0.4331334  P(node) =0.08987864
##     class counts:  3497  2672
##    probabilities: 0.567 0.433
##   left son=96 (5887 obs) right son=97 (282 obs)
##   Primary splits:
##       discharge_disposition_id < 10.5      to the left,  improve=19.22173, (0 missing)
##       number_diagnoses         < 6.5       to the right, improve=15.52569, (0 missing)
```

```
##       patient_nbr              < 42631400  to the left,  improve=14.29093, (0 missing)
##       diag_1                   < 237.5     to the right, improve=13.55089, (0 missing)
##       diag_3                   < 401.5     to the right, improve=12.27712, (0 missing)
##
## Node number 49: 5739 observations,    complexity param=0.003941301
##   predicted class=0  expected loss=0.49364  P(node) =0.08361379
##     class counts:  2833  2906
##    probabilities: 0.494 0.506
##   left son=98 (1110 obs) right son=99 (4629 obs)
##   Primary splits:
##       number_outpatient < 0.5       to the right, improve=21.04423, (0 missing)
##       number_emergency  < 0.5       to the right, improve=20.90247, (0 missing)
##       patient_nbr       < 83960980  to the right, improve=20.36068, (0 missing)
##       insulinSteady     < 0.5       to the left,  improve=19.24421, (0 missing)
##       age.80.90.        < 0.5       to the right, improve=11.12052, (0 missing)
##   Surrogate splits:
##       number_emergency < 4.5       to the right, agree=0.807, adj=0.002, (0 split)
##
## Node number 50: 223 observations
##   predicted class=1  expected loss=0.3632287  P(node) =0.003248976
##     class counts:   142    81
##    probabilities: 0.637 0.363
##
## Node number 51: 3672 observations,    complexity param=0.0005815886
##   predicted class=0  expected loss=0.4223856  P(node) =0.05349884
##     class counts:  1551  2121
##    probabilities: 0.422 0.578
##   left son=102 (1956 obs) right son=103 (1716 obs)
##   Primary splits:
##       encounter_id             < 151618900 to the left,  improve=15.739060, (0 missing)
##       discharge_disposition_id < 9.5       to the left,  improve=13.666500, (0 missing)
##       number_diagnoses         < 6.5       to the right, improve=11.663750, (0 missing)
##       num_medications          < 6.5       to the right, improve=10.558490, (0 missing)
##       number_outpatient        < 0.5       to the right, improve= 9.749345, (0 missing)
##   Surrogate splits:
##       number_diagnoses  < 8.5       to the left,  agree=0.558, adj=0.054, (0 split)
##       number_outpatient < 0.5       to the left,  agree=0.556, adj=0.049, (0 split)
##       patient_nbr       < 44314210  to the right, agree=0.550, adj=0.037, (0 split)
##       num_medications   < 20.5      to the left,  agree=0.547, adj=0.031, (0 split)
##       A1CresultNorm     < 0.5       to the left,  agree=0.546, adj=0.029, (0 split)
##
## Node number 52: 911 observations,    complexity param=0.000550432
##   predicted class=0  expected loss=0.4994512  P(node) =0.01327272
##     class counts:   455   456
##    probabilities: 0.499 0.501
##   left son=104 (549 obs) right son=105 (362 obs)
##   Primary splits:
##       encounter_id     < 289944600 to the left,  improve=6.585363, (0 missing)
##       num_medications  < 8.5       to the right, improve=5.427270, (0 missing)
##       diag_3           < 250.005   to the right, improve=4.923062, (0 missing)
##       patient_nbr      < 111962900 to the right, improve=4.580543, (0 missing)
##       number_diagnoses < 8.5       to the right, improve=4.536434, (0 missing)
##   Surrogate splits:
##       patient_nbr              < 44657550  to the right, agree=0.609, adj=0.017, (0 spl
it)
##       age.90.100.              < 0.5       to the left,  agree=0.609, adj=0.017, (0 spl
it)
##       num_medications          < 2.5       to the right, agree=0.608, adj=0.014, (0 spl
it)
```

```
##       discharge_disposition_id < 23.5      to the left,  agree=0.607, adj=0.011, (0 spl
it)
##       number_diagnoses         < 9.5       to the left,  agree=0.607, adj=0.011, (0 spl
it)
##
## Node number 53: 7584 observations
##   predicted class=0  expected loss=0.3420359  P(node) =0.1104943
##     class counts:  2594  4990
##    probabilities: 0.342 0.658
##
## Node number 56: 2071 observations,    complexity param=0.0004569624
##   predicted class=0  expected loss=0.4934814  P(node) =0.03017323
##     class counts:  1022  1049
##    probabilities: 0.493 0.507
##   left son=112 (2012 obs) right son=113 (59 obs)
##   Primary splits:
##       age.90.100.              < 0.5       to the left,  improve=6.952111, (0 missing)
##       number_outpatient        < 0.5       to the right, improve=6.151181, (0 missing)
##       discharge_disposition_id < 6.5       to the left,  improve=5.723952, (0 missing)
##       patient_nbr              < 1148324   to the left,  improve=5.227575, (0 missing)
##       num_medications          < 5.5       to the right, improve=5.016361, (0 missing)
##
## Node number 57: 1091 observations
##   predicted class=0  expected loss=0.3767186  P(node) =0.01589522
##     class counts:   411   680
##    probabilities: 0.377 0.623
##
## Node number 80: 4221 observations
##   predicted class=1  expected loss=0.3383085  P(node) =0.06149744
##     class counts:  2793  1428
##    probabilities: 0.662 0.338
##
## Node number 81: 2227 observations,    complexity param=0.0003946494
##   predicted class=1  expected loss=0.4126628  P(node) =0.03244606
##     class counts:  1308   919
##    probabilities: 0.587 0.413
##   left son=162 (1841 obs) right son=163 (386 obs)
##   Primary splits:
##       encounter_id        < 210339600 to the left,  improve=9.884537, (0 missing)
##       diag_3              < 424.5     to the right, improve=8.078673, (0 missing)
##       admission_source_id < 1.5       to the left,  improve=7.173921, (0 missing)
##       number_outpatient   < 0.5       to the right, improve=5.003349, (0 missing)
##       patient_nbr         < 42580620  to the left,  improve=4.251902, (0 missing)
##
## Node number 82: 1755 observations,    complexity param=0.001417622
##   predicted class=1  expected loss=0.4883191  P(node) =0.0255693
##     class counts:   898   857
##    probabilities: 0.512 0.488
##   left son=164 (1589 obs) right son=165 (166 obs)
##   Primary splits:
##       num_medications  < 8.5       to the right, improve=9.656930, (0 missing)
##       diag_1           < 731.5     to the left,  improve=9.047702, (0 missing)
##       number_emergency < 0.5       to the right, improve=8.505479, (0 missing)
##       raceHispanic     < 0.5       to the left,  improve=5.198239, (0 missing)
##       patient_nbr      < 94600120  to the left,  improve=4.512646, (0 missing)
##
## Node number 83: 525 observations
##   predicted class=0  expected loss=0.3790476  P(node) =0.007648936
##     class counts:   199   326
```

```
##    probabilities: 0.379 0.621
##
## Node number 86: 1091 observations,    complexity param=0.0009814307
##   predicted class=0  expected loss=0.492209  P(node) =0.01589522
##     class counts:   537   554
##    probabilities: 0.492 0.508
##   left son=172 (873 obs) right son=173 (218 obs)
##   Primary splits:
##       encounter_id       < 156110500 to the left,  improve=16.819650, (0 missing)
##       patient_nbr        < 24698760  to the left,  improve=14.758930, (0 missing)
##       num_lab_procedures < 21.5      to the right, improve= 9.542498, (0 missing)
##       admission_type_id  < 2.5       to the left,  improve= 5.468388, (0 missing)
##       number_emergency   < 0.5       to the right, improve= 4.802327, (0 missing)
##
## Node number 87: 982 observations,    complexity param=0.0003738784
##   predicted class=0  expected loss=0.3533605  P(node) =0.01430715
##     class counts:   347   635
##    probabilities: 0.353 0.647
##   left son=174 (114 obs) right son=175 (868 obs)
##   Primary splits:
##       encounter_id        < 119080400 to the right, improve=10.242720, (0 missing)
##       admission_source_id < 6.5       to the right, improve= 9.345943, (0 missing)
##       diag_2              < 562       to the right, improve= 8.837783, (0 missing)
##       admission_type_id   < 4.5       to the right, improve= 6.075962, (0 missing)
##       number_emergency    < 1.5       to the right, improve= 6.037780, (0 missing)
##   Surrogate splits:
##       number_emergency  < 0.5       to the right, agree=0.902, adj=0.158, (0 split)
##       number_outpatient < 1.5       to the right, agree=0.898, adj=0.123, (0 split)
##       admission_type_id < 4.5       to the right, agree=0.894, adj=0.088, (0 split)
##       max_glu_serumNone < 0.5       to the left,  agree=0.890, adj=0.053, (0 split)
##       max_glu_serumNorm < 0.5       to the right, agree=0.888, adj=0.035, (0 split)
##
## Node number 90: 293 observations
##   predicted class=1  expected loss=0.4197952  P(node) =0.004268835
##     class counts:   170   123
##    probabilities: 0.580 0.420
##
## Node number 91: 373 observations
##   predicted class=0  expected loss=0.4423592  P(node) =0.005434387
##     class counts:   165   208
##    probabilities: 0.442 0.558
##
## Node number 96: 5887 observations,    complexity param=0.001137213
##   predicted class=1  expected loss=0.4244946  P(node) =0.08577007
##     class counts:  3388  2499
##    probabilities: 0.576 0.424
##   left son=192 (3948 obs) right son=193 (1939 obs)
##   Primary splits:
##       number_diagnoses < 6.5       to the right, improve=16.92624, (0 missing)
##       diag_3           < 401.5     to the right, improve=15.65959, (0 missing)
##       age.50.60.       < 0.5       to the left,  improve=12.79176, (0 missing)
##       patient_nbr      < 42631400  to the left,  improve=12.41752, (0 missing)
##       age.70.80.       < 0.5       to the right, improve=11.92452, (0 missing)
##   Surrogate splits:
##       max_glu_serumNone < 0.5       to the right, agree=0.691, adj=0.061, (0 split)
##       num_medications   < 6.5       to the right, agree=0.685, adj=0.045, (0 split)
##       diag_2            < 250.015   to the right, agree=0.682, adj=0.033, (0 split)
##       diag_3            < 250.005   to the right, agree=0.678, adj=0.022, (0 split)
##       max_glu_serumNorm < 0.5       to the left,  agree=0.678, adj=0.021, (0 split)
```

```
##
## Node number 97: 282 observations,    complexity param=0.001339731
##   predicted class=0  expected loss=0.3865248  P(node) =0.004108571
##     class counts:   109   173
##    probabilities: 0.387 0.613
##   left son=194 (171 obs) right son=195 (111 obs)
##   Primary splits:
##       discharge_disposition_id < 14.5      to the right, improve=49.716090, (0 missing)
##       diag_1                   < 640.5     to the right, improve= 9.531637, (0 missing)
##       time_in_hospital         < 2.5       to the right, improve= 9.421426, (0 missing)
##       encounter_id             < 85356030  to the right, improve= 9.223700, (0 missing)
##       patient_nbr              < 43177550  to the left,  improve= 7.886071, (0 missing)
##   Surrogate splits:
##       encounter_id       < 65238040  to the right, agree=0.702, adj=0.243, (0 split)
##       time_in_hospital   < 2.5       to the right, agree=0.695, adj=0.225, (0 split)
##       num_lab_procedures < 60.5      to the left,  agree=0.688, adj=0.207, (0 split)
##       diag_1             < 433       to the right, agree=0.677, adj=0.180, (0 split)
##       diag_2             < 227       to the right, agree=0.667, adj=0.153, (0 split)
##
## Node number 98: 1110 observations,    complexity param=0.000467348
##   predicted class=1  expected loss=0.4189189  P(node) =0.01617204
##     class counts:   645   465
##    probabilities: 0.581 0.419
##   left son=196 (909 obs) right son=197 (201 obs)
##   Primary splits:
##       diag_3             < 250.015   to the right, improve=6.880948, (0 missing)
##       num_procedures     < 0.5       to the left,  improve=5.954425, (0 missing)
##       num_lab_procedures < 30.5      to the right, improve=5.447272, (0 missing)
##       age.30.40.         < 0.5       to the left,  improve=4.961895, (0 missing)
##       encounter_id       < 154363800 to the right, improve=4.712487, (0 missing)
##   Surrogate splits:
##       pioglitazoneDown < 0.5       to the left,  agree=0.821, adj=0.01, (0 split)
##
## Node number 99: 4629 observations,    complexity param=0.002710618
##   predicted class=0  expected loss=0.4726723  P(node) =0.06744176
##     class counts:  2188  2441
##    probabilities: 0.473 0.527
##   left son=198 (343 obs) right son=199 (4286 obs)
##   Primary splits:
##       number_emergency < 0.5       to the right, improve=17.605380, (0 missing)
##       insulinSteady    < 0.5       to the left,  improve=15.481100, (0 missing)
##       patient_nbr      < 83960980  to the right, improve=12.801370, (0 missing)
##       time_in_hospital < 1.5       to the right, improve= 6.913723, (0 missing)
##       age.80.90.       < 0.5       to the right, improve= 6.656896, (0 missing)
##
## Node number 102: 1956 observations,    complexity param=0.0005815886
##   predicted class=0  expected loss=0.4657464  P(node) =0.02849775
##     class counts:   911  1045
##    probabilities: 0.466 0.534
##   left son=204 (1015 obs) right son=205 (941 obs)
##   Primary splits:
##       number_diagnoses         < 7.5       to the right, improve=11.189350, (0 missing)
##       discharge_disposition_id < 9.5       to the left,  improve=10.451170, (0 missing)
##       num_medications          < 4.5       to the right, improve= 8.554995, (0 missing)
##       diag_1                   < 250.105   to the right, improve= 8.551314, (0 missing)
##       diag_3                   < 250.005   to the right, improve= 6.107894, (0 missing)
##   Surrogate splits:
##       admission_type_id < 2.5       to the left,  agree=0.625, adj=0.221, (0 split)
##       diag_3            < 401.5     to the right, agree=0.621, adj=0.211, (0 split)
```

```
##       num_medications   < 11.5      to the right, agree=0.602, adj=0.172, (0 split)
##       diag_2            < 401.5     to the right, agree=0.598, adj=0.165, (0 split)
##       time_in_hospital  < 2.5       to the right, agree=0.598, adj=0.164, (0 split)
##
## Node number 103: 1716 observations,    complexity param=0.0005815886
##   predicted class=0  expected loss=0.3729604  P(node) =0.02500109
##     class counts:   640  1076
##    probabilities: 0.373 0.627
##   left son=206 (127 obs) right son=207 (1589 obs)
##   Primary splits:
##       number_outpatient  < 1.5       to the right, improve=12.064060, (0 missing)
##       number_diagnoses   < 6.5       to the right, improve= 9.443069, (0 missing)
##       patient_nbr        < 82010930  to the right, improve= 9.369881, (0 missing)
##       num_lab_procedures < 38.5      to the right, improve= 6.812501, (0 missing)
##       age.30.40.         < 0.5       to the left,  improve= 5.015571, (0 missing)
##   Surrogate splits:
##       patient_nbr < 114492500 to the right, agree=0.927, adj=0.008, (0 split)
##
## Node number 104: 549 observations,    complexity param=0.0004361914
##   predicted class=1  expected loss=0.4517304  P(node) =0.007998601
##     class counts:   301   248
##    probabilities: 0.548 0.452
##   left son=208 (475 obs) right son=209 (74 obs)
##   Primary splits:
##       num_medications  < 8.5       to the right, improve=3.491300, (0 missing)
##       diag_3           < 250.005   to the right, improve=2.866612, (0 missing)
##       diabetesMedNo    < 0.5       to the left,  improve=2.662556, (0 missing)
##       diabetesMedYes   < 0.5       to the right, improve=2.662556, (0 missing)
##       number_diagnoses < 8.5       to the right, improve=2.661645, (0 missing)
##   Surrogate splits:
##       number_diagnoses < 3.5       to the right, agree=0.871, adj=0.041, (0 split)
##       raceAsian        < 0.5       to the left,  agree=0.867, adj=0.014, (0 split)
##
## Node number 105: 362 observations
##   predicted class=0  expected loss=0.4254144  P(node) =0.005274123
##     class counts:   154   208
##    probabilities: 0.425 0.575
##
## Node number 112: 2012 observations,    complexity param=0.0004569624
##   predicted class=1  expected loss=0.499503  P(node) =0.02931364
##     class counts:  1007  1005
##    probabilities: 0.500 0.500
##   left son=224 (93 obs) right son=225 (1919 obs)
##   Primary splits:
##       number_outpatient        < 0.5       to the right, improve=5.384797, (0 missing)
##       discharge_disposition_id < 6.5       to the left,  improve=5.262211, (0 missing)
##       num_medications          < 5.5       to the right, improve=5.076094, (0 missing)
##       patient_nbr              < 1148324   to the left,  improve=4.899006, (0 missing)
##       encounter_id             < 43168910  to the left,  improve=4.570270, (0 missing)
##   Surrogate splits:
##       number_emergency < 3.5       to the right, agree=0.954, adj=0.011, (0 split)
##
## Node number 113: 59 observations
##   predicted class=0  expected loss=0.2542373  P(node) =0.0008595947
##     class counts:    15    44
##    probabilities: 0.254 0.746
##
## Node number 162: 1841 observations
##   predicted class=1  expected loss=0.3910918  P(node) =0.02682227
```

```
##     class counts:  1121   720
##    probabilities: 0.609 0.391
##
## Node number 163: 386 observations,    complexity param=0.0003946494
##   predicted class=0  expected loss=0.484456  P(node) =0.005623789
##     class counts:   187   199
##    probabilities: 0.484 0.516
##   left son=326 (138 obs) right son=327 (248 obs)
##   Primary splits:
##       diag_1             < 412.5     to the left,  improve=5.174037, (0 missing)
##       num_lab_procedures < 27.5      to the right, improve=4.338184, (0 missing)
##       genderFemale       < 0.5       to the right, improve=4.110162, (0 missing)
##       genderMale         < 0.5       to the left,  improve=4.110162, (0 missing)
##       number_emergency   < 1.5       to the right, improve=4.035438, (0 missing)
##   Surrogate splits:
##       time_in_hospital   < 7.5       to the right, agree=0.679, adj=0.101, (0 split)
##       diag_2             < 247       to the left,  agree=0.676, adj=0.094, (0 split)
##       num_procedures     < 0.5       to the left,  agree=0.666, adj=0.065, (0 split)
##       num_lab_procedures < 70.5      to the right, agree=0.655, adj=0.036, (0 split)
##       admission_type_id  < 5.5       to the right, agree=0.650, adj=0.022, (0 split)
##
## Node number 164: 1589 observations,    complexity param=0.0006387089
##   predicted class=1  expected loss=0.4713656  P(node) =0.02315078
##     class counts:   840   749
##    probabilities: 0.529 0.471
##   left son=328 (365 obs) right son=329 (1224 obs)
##   Primary splits:
##       number_emergency < 0.5       to the right, improve=8.738109, (0 missing)
##       diag_1           < 731.5     to the left,  improve=7.811260, (0 missing)
##       num_medications  < 20.5      to the right, improve=4.346174, (0 missing)
##       raceHispanic     < 0.5       to the left,  improve=3.997236, (0 missing)
##       A1Cresult.7      < 0.5       to the right, improve=3.890047, (0 missing)
##   Surrogate splits:
##       admission_source_id < 8         to the right, agree=0.772, adj=0.005, (0 split)
##       number_outpatient   < 13        to the right, agree=0.772, adj=0.005, (0 split)
##       repaglinideUp       < 0.5       to the right, agree=0.772, adj=0.005, (0 split)
##       patient_nbr         < 28166140  to the left,  agree=0.771, adj=0.003, (0 split)
##       num_lab_procedures  < 88.5      to the right, agree=0.771, adj=0.003, (0 split)
##
## Node number 165: 166 observations
##   predicted class=0  expected loss=0.3493976  P(node) =0.002418521
##     class counts:    58   108
##    probabilities: 0.349 0.651
##
## Node number 172: 873 observations,    complexity param=0.0009814307
##   predicted class=1  expected loss=0.4639175  P(node) =0.01271909
##     class counts:   468   405
##    probabilities: 0.536 0.464
##   left son=344 (379 obs) right son=345 (494 obs)
##   Primary splits:
##       patient_nbr         < 23603180  to the left,  improve=17.910880, (0 missing)
##       num_lab_procedures  < 21.5      to the right, improve=14.894610, (0 missing)
##       raceCaucasian       < 0.5       to the left,  improve=10.846140, (0 missing)
##       raceAfricanAmerican < 0.5       to the right, improve= 9.445129, (0 missing)
##       admission_source_id < 12        to the left,  improve= 8.794277, (0 missing)
##   Surrogate splits:
##       num_lab_procedures < 52.5      to the right, agree=0.577, adj=0.026, (0 split)
##       age.20.30.         < 0.5       to the right, agree=0.576, adj=0.024, (0 split)
##       number_emergency   < 2.5       to the right, agree=0.575, adj=0.021, (0 split)
```

```
##       diag_1             < 796       to the right, agree=0.575, adj=0.021, (0 split)
##       age.80.90.         < 0.5       to the right, agree=0.574, adj=0.018, (0 split)
##
## Node number 173: 218 observations
##   predicted class=0  expected loss=0.3165138  P(node) =0.003176129
##     class counts:    69   149
##    probabilities: 0.317 0.683
##
## Node number 174: 114 observations
##   predicted class=1  expected loss=0.4473684  P(node) =0.001660912
##     class counts:    63    51
##    probabilities: 0.553 0.447
##
## Node number 175: 868 observations
##   predicted class=0  expected loss=0.3271889  P(node) =0.01264624
##     class counts:   284   584
##    probabilities: 0.327 0.673
##
## Node number 192: 3948 observations
##   predicted class=1  expected loss=0.397923  P(node) =0.05752
##     class counts:  2377  1571
##    probabilities: 0.602 0.398
##
## Node number 193: 1939 observations,    complexity param=0.001137213
##   predicted class=1  expected loss=0.4785972  P(node) =0.02825007
##     class counts:  1011   928
##    probabilities: 0.521 0.479
##   left son=386 (1376 obs) right son=387 (563 obs)
##   Primary splits:
##       num_procedures    < 1.5       to the left,  improve=11.799270, (0 missing)
##       patient_nbr       < 41805060  to the left,  improve=11.023200, (0 missing)
##       age.50.60.        < 0.5       to the left,  improve= 8.219781, (0 missing)
##       admission_type_id < 7         to the left,  improve= 6.723731, (0 missing)
##       age.80.90.        < 0.5       to the right, improve= 6.196334, (0 missing)
##   Surrogate splits:
##       num_medications     < 23.5      to the left,  agree=0.740, adj=0.103, (0 split)
##       admission_type_id   < 7         to the left,  agree=0.717, adj=0.027, (0 split)
##       diag_2              < 997.5     to the left,  agree=0.714, adj=0.014, (0 split)
##       diag_3              < 997.5     to the left,  agree=0.712, adj=0.007, (0 split)
##       admission_source_id < 18.5      to the left,  agree=0.711, adj=0.004, (0 split)
##
## Node number 194: 171 observations
##   predicted class=1  expected loss=0.374269  P(node) =0.002491368
##     class counts:   107    64
##    probabilities: 0.626 0.374
##
## Node number 195: 111 observations
##   predicted class=0  expected loss=0.01801802  P(node) =0.001617204
##     class counts:     2   109
##    probabilities: 0.018 0.982
##
## Node number 196: 909 observations
##   predicted class=1  expected loss=0.3927393  P(node) =0.01324359
##     class counts:   552   357
##    probabilities: 0.607 0.393
##
## Node number 197: 201 observations,    complexity param=0.0004050349
##   predicted class=0  expected loss=0.4626866  P(node) =0.00292845
##     class counts:    93   108
```

```
##    probabilities: 0.463 0.537
##   left son=394 (113 obs) right son=395 (88 obs)
##   Primary splits:
##       encounter_id      < 171786800 to the right, improve=4.642632, (0 missing)
##       number_diagnoses  < 6.5       to the right, improve=3.754252, (0 missing)
##       glyburideSteady   < 0.5       to the right, improve=3.149090, (0 missing)
##       glyburideNo       < 0.5       to the left,  improve=2.511269, (0 missing)
##       admission_type_id < 2.5       to the left,  improve=2.312521, (0 missing)
##   Surrogate splits:
##       metforminSteady          < 0.5       to the left,  agree=0.617, adj=0.125, (0 spl
it)
##       metforminNo              < 0.5       to the right, agree=0.612, adj=0.114, (0 spl
it)
##       age.40.50.               < 0.5       to the left,  agree=0.597, adj=0.080, (0 spl
it)
##       discharge_disposition_id < 5         to the left,  agree=0.597, adj=0.080, (0 spl
it)
##       patient_nbr              < 82497730  to the right, agree=0.587, adj=0.057, (0 spl
it)
##
## Node number 198: 343 observations
##   predicted class=1  expected loss=0.3731778  P(node) =0.004997305
##     class counts:   215   128
##    probabilities: 0.627 0.373
##
## Node number 199: 4286 observations,    complexity param=0.002274427
##   predicted class=0  expected loss=0.460336  P(node) =0.06244445
##     class counts:  1973  2313
##    probabilities: 0.460 0.540
##   left son=398 (527 obs) right son=399 (3759 obs)
##   Primary splits:
##       patient_nbr         < 43028900  to the left,  improve=14.258290, (0 missing)
##       insulinSteady       < 0.5       to the left,  improve=12.987860, (0 missing)
##       age.80.90.          < 0.5       to the right, improve= 8.332832, (0 missing)
##       admission_source_id < 8.5       to the left,  improve= 7.320104, (0 missing)
##       raceCaucasian       < 0.5       to the right, improve= 6.648617, (0 missing)
##   Surrogate splits:
##       num_medications < 57.5      to the right, agree=0.878, adj=0.004, (0 split)
##
## Node number 204: 1015 observations,    complexity param=0.0005815886
##   predicted class=1  expected loss=0.4827586  P(node) =0.01478794
##     class counts:   525   490
##    probabilities: 0.517 0.483
##   left son=408 (935 obs) right son=409 (80 obs)
##   Primary splits:
##       discharge_disposition_id < 9.5       to the left,  improve=7.280910, (0 missing)
##       number_outpatient        < 0.5       to the right, improve=5.224183, (0 missing)
##       encounter_id             < 137835100 to the right, improve=4.762192, (0 missing)
##       diag_1                   < 209.5     to the right, improve=3.735582, (0 missing)
##       num_medications          < 8.5       to the right, improve=3.236181, (0 missing)
##
## Node number 205: 941 observations
##   predicted class=0  expected loss=0.4102019  P(node) =0.01370981
##     class counts:   386   555
##    probabilities: 0.410 0.590
##
## Node number 206: 127 observations
##   predicted class=1  expected loss=0.4173228  P(node) =0.001850314
##     class counts:    74    53
```

```
##    probabilities: 0.583 0.417
##
## Node number 207: 1589 observations
##   predicted class=0  expected loss=0.3561989  P(node) =0.02315078
##     class counts:   566  1023
##    probabilities: 0.356 0.644
##
## Node number 208: 475 observations
##   predicted class=1  expected loss=0.4294737  P(node) =0.006920466
##     class counts:   271   204
##    probabilities: 0.571 0.429
##
## Node number 209: 74 observations
##   predicted class=0  expected loss=0.4054054  P(node) =0.001078136
##     class counts:    30    44
##    probabilities: 0.405 0.595
##
## Node number 224: 93 observations
##   predicted class=1  expected loss=0.3333333  P(node) =0.001354954
##     class counts:    62    31
##    probabilities: 0.667 0.333
##
## Node number 225: 1919 observations,    complexity param=0.0004569624
##   predicted class=0  expected loss=0.492444  P(node) =0.02795868
##     class counts:   945   974
##    probabilities: 0.492 0.508
##   left son=450 (413 obs) right son=451 (1506 obs)
##   Primary splits:
##       patient_nbr              < 1231384   to the left,  improve=6.169801, (0 missing)
##       time_in_hospital         < 2.5       to the right, improve=5.826608, (0 missing)
##       encounter_id             < 43168910  to the left,  improve=5.491311, (0 missing)
##       num_medications          < 5.5       to the right, improve=5.225591, (0 missing)
##       discharge_disposition_id < 6.5       to the left,  improve=4.160692, (0 missing)
##   Surrogate splits:
##       encounter_id < 3066159   to the left,  agree=0.785, adj=0.002, (0 split)
##
## Node number 326: 138 observations
##   predicted class=1  expected loss=0.4057971  P(node) =0.002010577
##     class counts:    82    56
##    probabilities: 0.594 0.406
##
## Node number 327: 248 observations
##   predicted class=0  expected loss=0.4233871  P(node) =0.003613212
##     class counts:   105   143
##    probabilities: 0.423 0.577
##
## Node number 328: 365 observations
##   predicted class=1  expected loss=0.3753425  P(node) =0.005317831
##     class counts:   228   137
##    probabilities: 0.625 0.375
##
## Node number 329: 1224 observations,    complexity param=0.0006387089
##   predicted class=1  expected loss=0.5  P(node) =0.01783295
##     class counts:   612   612
##    probabilities: 0.500 0.500
##   left son=658 (1081 obs) right son=659 (143 obs)
##   Primary splits:
##       diag_1          < 780.5     to the left,  improve=6.655143, (0 missing)
##       diag_2          < 399       to the right, improve=3.857875, (0 missing)
```

```
##       diag_3          < 32.5      to the right, improve=3.659801, (0 missing)
##       num_medications < 13.5      to the right, improve=3.591855, (0 missing)
##       raceHispanic    < 0.5       to the left,  improve=2.984504, (0 missing)
##   Surrogate splits:
##       diag_3 < 997.5     to the left,  agree=0.885, adj=0.014, (0 split)
##
## Node number 344: 379 observations
##   predicted class=1  expected loss=0.348285  P(node) =0.005521803
##     class counts:   247   132
##    probabilities: 0.652 0.348
##
## Node number 345: 494 observations,    complexity param=0.0004050349
##   predicted class=0  expected loss=0.4473684  P(node) =0.007197284
##     class counts:   221   273
##    probabilities: 0.447 0.553
##   left son=690 (448 obs) right son=691 (46 obs)
##   Primary splits:
##       patient_nbr         < 25150030  to the left,  improve=8.840021, (0 missing)
##       num_lab_procedures  < 17.5      to the right, improve=8.450326, (0 missing)
##       admission_type_id   < 2.5       to the left,  improve=4.618832, (0 missing)
##       diag_2              < 491.5     to the left,  improve=3.750297, (0 missing)
##       admission_source_id < 12        to the left,  improve=3.050947, (0 missing)
##   Surrogate splits:
##       num_lab_procedures < 58.5      to the left,  agree=0.913, adj=0.065, (0 split)
##
## Node number 386: 1376 observations
##   predicted class=1  expected loss=0.443314  P(node) =0.0200475
##     class counts:   766   610
##    probabilities: 0.557 0.443
##
## Node number 387: 563 observations,    complexity param=0.0003427218
##   predicted class=0  expected loss=0.4351687  P(node) =0.008202573
##     class counts:   245   318
##    probabilities: 0.435 0.565
##   left son=774 (452 obs) right son=775 (111 obs)
##   Primary splits:
##       encounter_id        < 139322200 to the left,  improve=5.256196, (0 missing)
##       diag_1              < 722.5     to the right, improve=3.971853, (0 missing)
##       age.50.60.          < 0.5       to the left,  improve=2.777564, (0 missing)
##       diag_2              < 690       to the right, improve=2.696662, (0 missing)
##       admission_source_id < 12        to the right, improve=2.674563, (0 missing)
##   Surrogate splits:
##       patient_nbr < 41335400  to the right, agree=0.805, adj=0.009, (0 split)
##
## Node number 394: 113 observations
##   predicted class=1  expected loss=0.4424779  P(node) =0.001646342
##     class counts:    63    50
##    probabilities: 0.558 0.442
##
## Node number 395: 88 observations
##   predicted class=0  expected loss=0.3409091  P(node) =0.001282107
##     class counts:    30    58
##    probabilities: 0.341 0.659
##
## Node number 398: 527 observations,    complexity param=0.0005919741
##   predicted class=1  expected loss=0.43074  P(node) =0.007678075
##     class counts:   300   227
##    probabilities: 0.569 0.431
##   left son=796 (438 obs) right son=797 (89 obs)
```

```
##   Primary splits:
##       A1CresultNone      < 0.5       to the right, improve=6.634213, (0 missing)
##       admission_type_id  < 1.5       to the left,  improve=5.203525, (0 missing)
##       pioglitazoneSteady < 0.5       to the right, improve=4.343874, (0 missing)
##       raceCaucasian      < 0.5       to the right, improve=3.542062, (0 missing)
##       pioglitazoneNo     < 0.5       to the left,  improve=3.324748, (0 missing)
##   Surrogate splits:
##       A1Cresult.8        < 0.5       to the left,  agree=0.894, adj=0.371, (0 split)
##       A1Cresult.7        < 0.5       to the left,  agree=0.890, adj=0.348, (0 split)
##       A1CresultNorm      < 0.5       to the left,  agree=0.879, adj=0.281, (0 split)
##       num_lab_procedures < 80.5      to the left,  agree=0.839, adj=0.045, (0 split)
##       raceAsian          < 0.5       to the left,  agree=0.837, adj=0.034, (0 split)
##
## Node number 399: 3759 observations,    complexity param=0.001604561
##   predicted class=0  expected loss=0.4450652  P(node) =0.05476638
##     class counts:  1673  2086
##    probabilities: 0.445 0.555
##   left son=798 (1824 obs) right son=799 (1935 obs)
##   Primary splits:
##       patient_nbr              < 82361890  to the right, improve=20.125150, (0 missing)
##       insulinSteady            < 0.5       to the left,  improve= 8.831511, (0 missing)
##       number_diagnoses         < 6.5       to the right, improve= 7.445804, (0 missing)
##       discharge_disposition_id < 6.5       to the left,  improve= 7.335428, (0 missing)
##       age.80.90.               < 0.5       to the right, improve= 7.106943, (0 missing)
##   Surrogate splits:
##       insulinNo          < 0.5       to the right, agree=0.568, adj=0.110, (0 split)
##       insulinSteady      < 0.5       to the left,  agree=0.565, adj=0.104, (0 split)
##       encounter_id       < 162975100 to the left,  agree=0.559, adj=0.091, (0 split)
##       admission_type_id  < 1.5       to the right, agree=0.554, adj=0.080, (0 split)
##       num_lab_procedures < 45.5      to the right, agree=0.537, adj=0.045, (0 split)
##
## Node number 408: 935 observations
##   predicted class=1  expected loss=0.4652406  P(node) =0.01362239
##     class counts:   500   435
##    probabilities: 0.535 0.465
##
## Node number 409: 80 observations
##   predicted class=0  expected loss=0.3125  P(node) =0.001165552
##     class counts:    25    55
##    probabilities: 0.312 0.688
##
## Node number 450: 413 observations,    complexity param=0.0003738784
##   predicted class=1  expected loss=0.4309927  P(node) =0.006017163
##     class counts:   235   178
##    probabilities: 0.569 0.431
##   left son=900 (319 obs) right son=901 (94 obs)
##   Primary splits:
##       time_in_hospital    < 2.5       to the right, improve=4.294926, (0 missing)
##       num_medications     < 8.5       to the right, improve=3.639601, (0 missing)
##       num_lab_procedures  < 65.5      to the right, improve=2.833016, (0 missing)
##       diag_3              < 427.5     to the right, improve=2.796881, (0 missing)
##       rosiglitazoneSteady < 0.5       to the right, improve=2.518227, (0 missing)
##   Surrogate splits:
##       num_medications    < 6.5       to the right, agree=0.801, adj=0.128, (0 split)
##       pioglitazoneSteady < 0.5       to the left,  agree=0.777, adj=0.021, (0 split)
##       num_lab_procedures < 26        to the right, agree=0.775, adj=0.011, (0 split)
##       repaglinideSteady  < 0.5       to the left,  agree=0.775, adj=0.011, (0 split)
##       pioglitazoneNo     < 0.5       to the right, agree=0.775, adj=0.011, (0 split)
##
```

```
## Node number 451: 1506 observations,    complexity param=0.0004569624
##   predicted class=0  expected loss=0.4714475  P(node) =0.02194152
##     class counts:   710   796
##    probabilities: 0.471 0.529
##   left son=902 (399 obs) right son=903 (1107 obs)
##   Primary splits:
##       encounter_id             < 24159970  to the left,  improve=4.931676, (0 missing)
##       diag_1                   < 46.5      to the right, improve=4.888764, (0 missing)
##       discharge_disposition_id < 6.5       to the left,  improve=4.529396, (0 missing)
##       num_medications          < 5.5       to the right, improve=4.310934, (0 missing)
##       diag_2                   < 250.005   to the right, improve=4.174118, (0 missing)
##   Surrogate splits:
##       discharge_disposition_id < 21.5      to the right, agree=0.780, adj=0.168, (0 spl
it)
##       admission_type_id        < 5.5       to the right, agree=0.761, adj=0.098, (0 spl
it)
##       admission_source_id      < 18.5      to the right, agree=0.752, adj=0.065, (0 spl
it)
##       glimepirideDown          < 0.5       to the right, agree=0.737, adj=0.008, (0 spl
it)
##       troglitazoneNo           < 0.5       to the left,  agree=0.736, adj=0.005, (0 spl
it)
##
## Node number 658: 1081 observations,    complexity param=0.0005192755
##   predicted class=1  expected loss=0.4810361  P(node) =0.01574952
##     class counts:   561   520
##    probabilities: 0.519 0.481
##   left son=1316 (1062 obs) right son=1317 (19 obs)
##   Primary splits:
##       diag_3          < 32.5      to the right, improve=3.679759, (0 missing)
##       num_medications < 22.5      to the right, improve=3.588078, (0 missing)
##       patient_nbr     < 82196470  to the right, improve=3.399767, (0 missing)
##       diag_2          < 407       to the right, improve=2.900157, (0 missing)
##       num_procedures  < 0.5       to the left,  improve=2.274520, (0 missing)
##
## Node number 659: 143 observations
##   predicted class=0  expected loss=0.3566434  P(node) =0.002083424
##     class counts:    51    92
##    probabilities: 0.357 0.643
##
## Node number 690: 448 observations,    complexity param=0.0004050349
##   predicted class=0  expected loss=0.4776786  P(node) =0.006527092
##     class counts:   214   234
##    probabilities: 0.478 0.522
##   left son=1380 (348 obs) right son=1381 (100 obs)
##   Primary splits:
##       num_lab_procedures < 21.5      to the right, improve=5.615181, (0 missing)
##       diag_2             < 511.5     to the left,  improve=4.709719, (0 missing)
##       number_emergency   < 0.5       to the right, improve=2.841058, (0 missing)
##       admission_type_id  < 2.5       to the left,  improve=2.637435, (0 missing)
##       glipizideSteady    < 0.5       to the right, improve=2.590030, (0 missing)
##   Surrogate splits:
##       admission_source_id < 12        to the left,  agree=0.946, adj=0.76, (0 split)
##       max_glu_serumNone   < 0.5       to the right, agree=0.938, adj=0.72, (0 split)
##       admission_type_id   < 4         to the left,  agree=0.913, adj=0.61, (0 split)
##       max_glu_serumNorm   < 0.5       to the left,  agree=0.873, adj=0.43, (0 split)
##       encounter_id        < 95473190  to the right, agree=0.844, adj=0.30, (0 split)
##
## Node number 691: 46 observations
```

```
##   predicted class=0  expected loss=0.1521739  P(node) =0.0006701925
##     class counts:     7    39
##    probabilities: 0.152 0.848
##
## Node number 774: 452 observations,    complexity param=0.0003427218
##   predicted class=0  expected loss=0.4690265  P(node) =0.006585369
##     class counts:   212   240
##    probabilities: 0.469 0.531
##   left son=1548 (60 obs) right son=1549 (392 obs)
##   Primary splits:
##       diag_1              < 723.5     to the right, improve=6.354852, (0 missing)
##       admission_source_id < 6.5       to the right, improve=3.296570, (0 missing)
##       number_diagnoses    < 3.5       to the right, improve=2.861376, (0 missing)
##       age.50.60.          < 0.5       to the left,  improve=2.696380, (0 missing)
##       diag_2              < 694       to the right, improve=2.438660, (0 missing)
##   Surrogate splits:
##       diag_3 < 789       to the right, agree=0.869, adj=0.017, (0 split)
##
## Node number 775: 111 observations
##   predicted class=0  expected loss=0.2972973  P(node) =0.001617204
##     class counts:    33    78
##    probabilities: 0.297 0.703
##
## Node number 796: 438 observations
##   predicted class=1  expected loss=0.3949772  P(node) =0.006381398
##     class counts:   265   173
##    probabilities: 0.605 0.395
##
## Node number 797: 89 observations
##   predicted class=0  expected loss=0.3932584  P(node) =0.001296677
##     class counts:    35    54
##    probabilities: 0.393 0.607
##
## Node number 798: 1824 observations,    complexity param=0.001604561
##   predicted class=0  expected loss=0.4983553  P(node) =0.02657459
##     class counts:   909   915
##    probabilities: 0.498 0.502
##   left son=1596 (561 obs) right son=1597 (1263 obs)
##   Primary splits:
##       patient_nbr              < 90616520  to the left,  improve=14.149050, (0 missing)
##       number_diagnoses         < 4.5       to the right, improve= 9.725327, (0 missing)
##       discharge_disposition_id < 6.5       to the left,  improve= 5.929306, (0 missing)
##       num_medications          < 35.5      to the left,  improve= 5.308161, (0 missing)
##       admission_source_id      < 8.5       to the left,  improve= 4.665676, (0 missing)
##   Surrogate splits:
##       max_glu_serum.300 < 0.5       to the right, agree=0.694, adj=0.005, (0 split)
##       encounter_id      < 213897100 to the right, agree=0.694, adj=0.004, (0 split)
##
## Node number 799: 1935 observations,    complexity param=0.0003349327
##   predicted class=0  expected loss=0.394832  P(node) =0.02819179
##     class counts:   764  1171
##    probabilities: 0.395 0.605
##   left son=1598 (767 obs) right son=1599 (1168 obs)
##   Primary splits:
##       patient_nbr              < 58758040  to the left,  improve=8.425702, (0 missing)
##       discharge_disposition_id < 2.5       to the right, improve=6.461318, (0 missing)
##       age.80.90.               < 0.5       to the right, improve=5.434772, (0 missing)
##       encounter_id             < 175135500 to the left,  improve=4.814331, (0 missing)
##       num_medications          < 15.5      to the right, improve=4.331392, (0 missing)
```

```
##   Surrogate splits:
##       num_medications          < 43.5      to the right, agree=0.609, adj=0.014, (0 spl
it)
##       num_lab_procedures       < 80.5      to the right, agree=0.608, adj=0.012, (0 spl
it)
##       encounter_id             < 154749100 to the left,  agree=0.606, adj=0.007, (0 spl
it)
##       discharge_disposition_id < 25.5      to the right, agree=0.606, adj=0.007, (0 spl
it)
##       nateglinideSteady        < 0.5       to the right, agree=0.606, adj=0.007, (0 spl
it)
##
## Node number 900: 319 observations
##   predicted class=1  expected loss=0.3918495  P(node) =0.004647639
##     class counts:   194   125
##    probabilities: 0.608 0.392
##
## Node number 901: 94 observations
##   predicted class=0  expected loss=0.4361702  P(node) =0.001369524
##     class counts:    41    53
##    probabilities: 0.436 0.564
##
## Node number 902: 399 observations
##   predicted class=1  expected loss=0.4611529  P(node) =0.005813191
##     class counts:   215   184
##    probabilities: 0.539 0.461
##
## Node number 903: 1107 observations
##   predicted class=0  expected loss=0.4471545  P(node) =0.01612833
##     class counts:   495   612
##    probabilities: 0.447 0.553
##
## Node number 1316: 1062 observations,    complexity param=0.0005192755
##   predicted class=1  expected loss=0.4755179  P(node) =0.0154727
##     class counts:   557   505
##    probabilities: 0.524 0.476
##   left son=2632 (221 obs) right son=2633 (841 obs)
##   Primary splits:
##       num_medications   < 22.5      to the right, improve=3.337510, (0 missing)
##       diag_2            < 399       to the right, improve=3.260434, (0 missing)
##       patient_nbr       < 82196470  to the right, improve=2.922899, (0 missing)
##       num_procedures    < 0.5       to the left,  improve=2.610984, (0 missing)
##       number_outpatient < 5.5       to the right, improve=1.936309, (0 missing)
##   Surrogate splits:
##       time_in_hospital   < 8.5       to the right, agree=0.812, adj=0.095, (0 split)
##       num_lab_procedures < 74.5      to the right, agree=0.796, adj=0.018, (0 split)
##       raceAsian          < 0.5       to the right, agree=0.793, adj=0.005, (0 split)
##       diag_2             < 997.5     to the right, agree=0.793, adj=0.005, (0 split)
##       diag_3             < 39.5      to the left,  agree=0.793, adj=0.005, (0 split)
##
## Node number 1317: 19 observations
##   predicted class=0  expected loss=0.2105263  P(node) =0.0002768186
##     class counts:     4    15
##    probabilities: 0.211 0.789
##
## Node number 1380: 348 observations,    complexity param=0.0004050349
##   predicted class=1  expected loss=0.4798851  P(node) =0.005070152
##     class counts:   181   167
##    probabilities: 0.520 0.480
```

```
##   left son=2760 (191 obs) right son=2761 (157 obs)
##   Primary splits:
##       patient_nbr         < 24428160  to the left,  improve=5.690512, (0 missing)
##       number_emergency    < 0.5       to the right, improve=5.560867, (0 missing)
##       encounter_id        < 92004980  to the left,  improve=5.162835, (0 missing)
##       admission_source_id < 12        to the right, improve=3.765238, (0 missing)
##       num_lab_procedures  < 25.5      to the left,  improve=3.765238, (0 missing)
##   Surrogate splits:
##       diag_1             < 572.5     to the left,  agree=0.598, adj=0.108, (0 split)
##       num_procedures     < 2.5       to the left,  agree=0.583, adj=0.076, (0 split)
##       diag_3             < 681.5     to the left,  agree=0.569, adj=0.045, (0 split)
##       num_lab_procedures < 50.5      to the left,  agree=0.566, adj=0.038, (0 split)
##       insulinSteady      < 0.5       to the left,  agree=0.566, adj=0.038, (0 split)
##
## Node number 1381: 100 observations
##   predicted class=0  expected loss=0.33  P(node) =0.00145694
##     class counts:    33    67
##    probabilities: 0.330 0.670
##
## Node number 1548: 60 observations
##   predicted class=1  expected loss=0.3166667  P(node) =0.0008741641
##     class counts:    41    19
##    probabilities: 0.683 0.317
##
## Node number 1549: 392 observations
##   predicted class=0  expected loss=0.4362245  P(node) =0.005711205
##     class counts:   171   221
##    probabilities: 0.436 0.564
##
## Node number 1596: 561 observations,    complexity param=0.0003738784
##   predicted class=1  expected loss=0.4081996  P(node) =0.008173434
##     class counts:   332   229
##    probabilities: 0.592 0.408
##   left son=3192 (378 obs) right son=3193 (183 obs)
##   Primary splits:
##       diag_3                   < 284.5     to the right, improve=5.431587, (0 missing)
##       diag_1                   < 494.5     to the left,  improve=4.107767, (0 missing)
##       num_medications          < 7.5       to the right, improve=3.866056, (0 missing)
##       discharge_disposition_id < 3.5       to the left,  improve=3.291135, (0 missing)
##       diag_2                   < 726       to the left,  improve=3.187697, (0 missing)
##   Surrogate splits:
##       number_diagnoses < 5.5       to the right, agree=0.688, adj=0.044, (0 split)
##       diag_2           < 41.5      to the right, agree=0.679, adj=0.016, (0 split)
##       encounter_id     < 214492300 to the left,  agree=0.677, adj=0.011, (0 split)
##       patient_nbr      < 82478560  to the right, agree=0.677, adj=0.011, (0 split)
##       raceHispanic     < 0.5       to the left,  agree=0.677, adj=0.011, (0 split)
##
## Node number 1597: 1263 observations
##   predicted class=0  expected loss=0.4568488  P(node) =0.01840115
##     class counts:   577   686
##    probabilities: 0.457 0.543
##
## Node number 1598: 767 observations,    complexity param=0.0003349327
##   predicted class=0  expected loss=0.452412  P(node) =0.01117473
##     class counts:   347   420
##    probabilities: 0.452 0.548
##   left son=3196 (392 obs) right son=3197 (375 obs)
##   Primary splits:
##       num_medications          < 15.5      to the right, improve=4.031178, (0 missing)
```

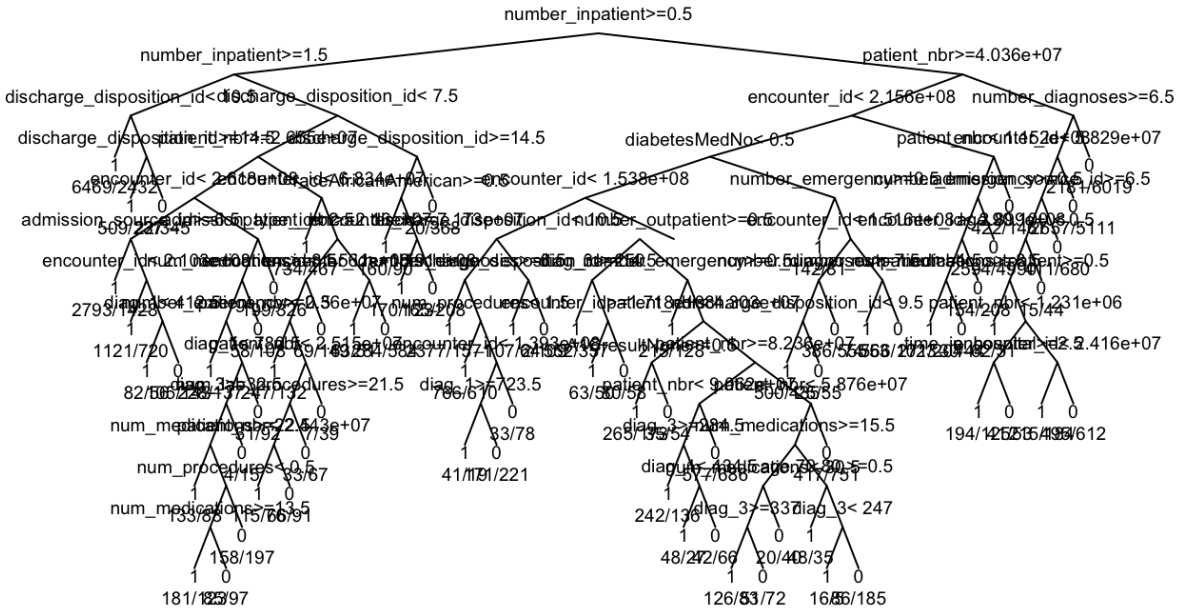
```
##       diag_2                   < 870.5     to the right, improve=3.898446, (0 missing)
##       age.20.30.               < 0.5       to the left,  improve=3.728993, (0 missing)
##       discharge_disposition_id < 18        to the right, improve=3.376231, (0 missing)
##       age.70.80.               < 0.5       to the right, improve=3.165855, (0 missing)
##   Surrogate splits:
##       num_procedures           < 0.5       to the right, agree=0.684, adj=0.355, (0 spl
it)
##       time_in_hospital         < 4.5       to the right, agree=0.675, adj=0.336, (0 spl
it)
##       num_lab_procedures       < 40.5      to the right, agree=0.652, adj=0.288, (0 spl
it)
##       number_diagnoses         < 8.5       to the right, agree=0.639, adj=0.261, (0 spl
it)
##       discharge_disposition_id < 1.5       to the right, agree=0.614, adj=0.211, (0 spl
it)
##
## Node number 1599: 1168 observations
##   predicted class=0  expected loss=0.3570205  P(node) =0.01701706
##     class counts:   417   751
##    probabilities: 0.357 0.643
##
## Node number 2632: 221 observations
##   predicted class=1  expected loss=0.39819  P(node) =0.003219838
##     class counts:   133    88
##    probabilities: 0.602 0.398
##
## Node number 2633: 841 observations,    complexity param=0.0005192755
##   predicted class=1  expected loss=0.4958383  P(node) =0.01225287
##     class counts:   424   417
##    probabilities: 0.504 0.496
##   left son=5266 (486 obs) right son=5267 (355 obs)
##   Primary splits:
##       num_procedures    < 0.5       to the left,  improve=4.290076, (0 missing)
##       genderFemale      < 0.5       to the left,  improve=2.846693, (0 missing)
##       genderMale        < 0.5       to the right, improve=2.846693, (0 missing)
##       patient_nbr       < 103621500 to the left,  improve=2.709852, (0 missing)
##       number_outpatient < 5.5       to the right, improve=2.198141, (0 missing)
##   Surrogate splits:
##       time_in_hospital    < 6.5       to the left,  agree=0.598, adj=0.048, (0 split)
##       admission_source_id < 4.5       to the right, agree=0.593, adj=0.037, (0 split)
##       num_lab_procedures  < 64.5      to the left,  agree=0.592, adj=0.034, (0 split)
##       num_medications     < 19.5      to the left,  agree=0.589, adj=0.025, (0 split)
##       diag_2              < 197.5     to the right, agree=0.585, adj=0.017, (0 split)
##
## Node number 2760: 191 observations
##   predicted class=1  expected loss=0.3979058  P(node) =0.002782756
##     class counts:   115    76
##    probabilities: 0.602 0.398
##
## Node number 2761: 157 observations
##   predicted class=0  expected loss=0.4203822  P(node) =0.002287396
##     class counts:    66    91
##    probabilities: 0.420 0.580
##
## Node number 3192: 378 observations
##   predicted class=1  expected loss=0.3597884  P(node) =0.005507234
##     class counts:   242   136
##    probabilities: 0.640 0.360
##
```

```
## Node number 3193: 183 observations,    complexity param=0.0003738784
##   predicted class=0  expected loss=0.4918033  P(node) =0.0026662
##     class counts:    90    93
##    probabilities: 0.492 0.508
##   left son=6386 (75 obs) right son=6387 (108 obs)
##   Primary splits:
##       diag_1       < 434.5     to the left,  improve=5.582077, (0 missing)
##       age.70.80.   < 0.5       to the right, improve=4.520466, (0 missing)
##       diag_2       < 601       to the left,  improve=4.129991, (0 missing)
##       genderFemale < 0.5       to the right, improve=2.720145, (0 missing)
##       genderMale   < 0.5       to the left,  improve=2.720145, (0 missing)
##   Surrogate splits:
##       raceAfricanAmerican < 0.5       to the right, agree=0.634, adj=0.107, (0 split)
##       num_procedures      < 4.5       to the right, agree=0.628, adj=0.093, (0 split)
##       raceCaucasian       < 0.5       to the left,  agree=0.617, adj=0.067, (0 split)
##       A1CresultNone       < 0.5       to the left,  agree=0.617, adj=0.067, (0 split)
##       A1CresultNorm       < 0.5       to the right, agree=0.617, adj=0.067, (0 split)
##
## Node number 3196: 392 observations,    complexity param=0.0003349327
##   predicted class=1  expected loss=0.497449  P(node) =0.005711205
##     class counts:   197   195
##    probabilities: 0.503 0.497
##   left son=6392 (332 obs) right son=6393 (60 obs)
##   Primary splits:
##       num_medications    < 30.5      to the left,  improve=4.057147, (0 missing)
##       num_lab_procedures < 58.5      to the left,  improve=3.683811, (0 missing)
##       diag_3             < 329       to the right, improve=3.170111, (0 missing)
##       insulinUp          < 0.5       to the left,  improve=3.039054, (0 missing)
##       encounter_id       < 183581200 to the left,  improve=2.766519, (0 missing)
##   Surrogate splits:
##       diag_1 < 996.5     to the left,  agree=0.849, adj=0.017, (0 split)
##       diag_3 < 962       to the left,  agree=0.849, adj=0.017, (0 split)
##
## Node number 3197: 375 observations,    complexity param=0.0003349327
##   predicted class=0  expected loss=0.4  P(node) =0.005463526
##     class counts:   150   225
##    probabilities: 0.400 0.600
##   left son=6394 (83 obs) right son=6395 (292 obs)
##   Primary splits:
##       age.70.80.               < 0.5       to the right, improve=6.778346, (0 missing)
##       diag_3                   < 237.5     to the left,  improve=5.977769, (0 missing)
##       encounter_id             < 156516700 to the left,  improve=4.564430, (0 missing)
##       diag_2                   < 802       to the right, improve=3.963536, (0 missing)
##       discharge_disposition_id < 5         to the right, improve=3.403258, (0 missing)
##   Surrogate splits:
##       time_in_hospital < 10        to the right, agree=0.784, adj=0.024, (0 split)
##       diag_2           < 936       to the right, agree=0.784, adj=0.024, (0 split)
##       diag_3           < 14.5      to the left,  agree=0.781, adj=0.012, (0 split)
##
## Node number 5266: 486 observations,    complexity param=0.0003738784
##   predicted class=1  expected loss=0.4526749  P(node) =0.007080729
##     class counts:   266   220
##    probabilities: 0.547 0.453
##   left son=10532 (304 obs) right son=10533 (182 obs)
##   Primary splits:
##       num_medications   < 13.5      to the right, improve=3.751544, (0 missing)
##       number_outpatient < 3.5       to the right, improve=1.985438, (0 missing)
##       diag_3            < 766.5     to the left,  improve=1.951934, (0 missing)
##       metforminNo       < 0.5       to the right, improve=1.728011, (0 missing)
```

```
##       patient_nbr       < 31885070  to the left,  improve=1.664886, (0 missing)
##   Surrogate splits:
##       time_in_hospital    < 1.5       to the right, agree=0.658, adj=0.088, (0 split)
##       raceCaucasian       < 0.5       to the right, agree=0.654, adj=0.077, (0 split)
##       raceAfricanAmerican < 0.5       to the left,  agree=0.652, adj=0.071, (0 split)
##       num_lab_procedures  < 26.5      to the right, agree=0.648, adj=0.060, (0 split)
##       number_diagnoses    < 6.5       to the right, agree=0.638, adj=0.033, (0 split)
##
## Node number 5267: 355 observations
##   predicted class=0  expected loss=0.4450704  P(node) =0.005172137
##     class counts:   158   197
##    probabilities: 0.445 0.555
##
## Node number 6386: 75 observations
##   predicted class=1  expected loss=0.36  P(node) =0.001092705
##     class counts:    48    27
##    probabilities: 0.640 0.360
##
## Node number 6387: 108 observations
##   predicted class=0  expected loss=0.3888889  P(node) =0.001573495
##     class counts:    42    66
##    probabilities: 0.389 0.611
##
## Node number 6392: 332 observations,    complexity param=0.0003349327
##   predicted class=1  expected loss=0.4668675  P(node) =0.004837041
##     class counts:   177   155
##    probabilities: 0.533 0.467
##   left son=12784 (209 obs) right son=12785 (123 obs)
##   Primary splits:
##       diag_3         < 337       to the right, improve=5.487212, (0 missing)
##       num_procedures < 4.5       to the right, improve=2.887451, (0 missing)
##       encounter_id   < 183525200 to the left,  improve=2.772205, (0 missing)
##       diag_1         < 57.5      to the right, improve=2.047181, (0 missing)
##       patient_nbr    < 45678780  to the right, improve=2.031936, (0 missing)
##   Surrogate splits:
##       number_diagnoses  < 4.5       to the right, agree=0.654, adj=0.065, (0 split)
##       encounter_id      < 207473600 to the left,  agree=0.651, adj=0.057, (0 split)
##       patient_nbr       < 43241090  to the right, agree=0.642, adj=0.033, (0 split)
##       age.40.50.        < 0.5       to the left,  agree=0.639, adj=0.024, (0 split)
##       admission_type_id < 5.5       to the left,  agree=0.636, adj=0.016, (0 split)
##
## Node number 6393: 60 observations
##   predicted class=0  expected loss=0.3333333  P(node) =0.0008741641
##     class counts:    20    40
##    probabilities: 0.333 0.667
##
## Node number 6394: 83 observations
##   predicted class=1  expected loss=0.4216867  P(node) =0.00120926
##     class counts:    48    35
##    probabilities: 0.578 0.422
##
## Node number 6395: 292 observations,    complexity param=0.0003349327
##   predicted class=0  expected loss=0.3493151  P(node) =0.004254265
##     class counts:   102   190
##    probabilities: 0.349 0.651
##   left son=12790 (21 obs) right son=12791 (271 obs)
##   Primary splits:
##       diag_3           < 247       to the left,  improve=7.703704, (0 missing)
##       age.40.50.       < 0.5       to the right, improve=4.397894, (0 missing)
```

```
##       encounter_id     < 156516700 to the left,  improve=3.436355, (0 missing)
##       time_in_hospital < 6.5       to the right, improve=3.200475, (0 missing)
##       glyburideNo      < 0.5       to the right, improve=1.882034, (0 missing)
##
## Node number 10532: 304 observations
##   predicted class=1  expected loss=0.4046053  P(node) =0.004429098
##     class counts:   181   123
##    probabilities: 0.595 0.405
##
## Node number 10533: 182 observations
##   predicted class=0  expected loss=0.467033  P(node) =0.002651631
##     class counts:    85    97
##    probabilities: 0.467 0.533
##
## Node number 12784: 209 observations
##   predicted class=1  expected loss=0.3971292  P(node) =0.003045005
##     class counts:   126    83
##    probabilities: 0.603 0.397
##
## Node number 12785: 123 observations
##   predicted class=0  expected loss=0.4146341  P(node) =0.001792036
##     class counts:    51    72
##    probabilities: 0.415 0.585
##
## Node number 12790: 21 observations
##   predicted class=1  expected loss=0.2380952  P(node) =0.0003059574
##     class counts:    16     5
##    probabilities: 0.762 0.238
##
## Node number 12791: 271 observations
##   predicted class=0  expected loss=0.3173432  P(node) =0.003948308
##     class counts:    86   185
##    probabilities: 0.317 0.683
```

```
plot(x_pruned, uniform = TRUE, compress = TRUE, branch = 0)
text(x_pruned, use.n = TRUE, cex = 0.6, xpd = NA)
```



#From the plot, we can see that the tree is deep enough, the criterias to deciding readmitted status are detail, which is good.

#3. Predict values for test data and compare confusion matrix with train. Type = "class" (R), and predict (Python) Are results stable?

```
x_test=rpart(readmitted~., data=test,control=rpart.control(cp=0.0003349327,minsplit=30,xval=10))
Prediction_Model_test<-predict(x_pruned, data=test, type="class")
confusionMatrix(Prediction_Model_test, train$readmitted)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction     1     0
##          1 19329 10531
##          0 12767 26010
##
##                Accuracy : 0.6606
##                  95% CI : (0.657, 0.6641)
##     No Information Rate : 0.5324
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.3154
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.6022
##             Specificity : 0.7118
##          Pos Pred Value : 0.6473
##          Neg Pred Value : 0.6708
##              Prevalence : 0.4676
##          Detection Rate : 0.2816
##    Detection Prevalence : 0.4350
##       Balanced Accuracy : 0.6570
##
##        'Positive' Class : 1
##
```

```
round(prop.table(table(Prediction_Model_test,train$readmitted),1),2)
```

```
##
## Prediction_Model_test    1    0
##                     1 0.65 0.35
##                     0 0.33 0.67
```

```
#The results of confusion matrix between train and test dataset are the same (stable)

# 4. Compare Test  results for logistic regression and classification tree by comparing th
e two confusion matrices.
#Answer: Proportion of TP and TN in classification tree is higher (> 50%) compared to logi
stic regression, and the pattern of interactions from confusion matrix in classification t
ree is easier to find.
```