HOSPITAL PATIENT READMISSIONS WITHIN 30 DAYS

SECTION 1 - OVERVIEW

Using data from a major U.S. hospital system that included 32 key indicators on 11,863 anonymous patients, a model was created to describe and predict the probability of 1,117 patients being readmitted within 30 days of leaving the hospital. The model was based on an estimate of the relative highest quality of the model using a total of 19 key indicators across patient demographics, vital signs, medical history, hospital stay, and healthcare utilization categories.

SECTION 2 - TECHNICAL DISCUSSION

A logistic regression model was developed leveraging the hospital data loading the code and data file into an R studio session. The data file contained the input data that was utilized by the regression model and the code file contained the functions to analyze the results of the model.

The logistic regression model utilized the generalized linear model ("glm") function within R and the option within this function was set at "family=binomial" to ensure the model was based on binary dependent variables. The data was defined within the regression model by setting data=hospital. The glm function regressed the dependent variable identified as "readmit30" against all the independent variables in the hospital data. The independent and dependent variables were reviewed to become familiar with the dataset for which there was 11,863 rows of data. Please note, each row within the data file represents a patient. The glm regression function was run and the regression was defined as the "fit" object. The fit object was loaded into the custom function "aa_critique_fit" which was used to evaluate the model.

The next step evaluated the results of the initial (first) logistic regression model. It was initially critiqued by reviewing the AIC value and the p-values for the model. These values were calculated by applying the summary function within R against the model. Per the summary output, the AIC value was 11,135. Also, based on a review of p-value results in the summary output, we noted several independent variables were labeled as insignificant.

There are 32 independent variables and the glm function created dummy variables, resulting in over 50 variables in total. The concern of having a high number of variables is the regression model can overfit the actual distribution. The AIC measures the goodness-of-fit of models and penalizes for increasing the number of variables to reduce the risk of overfitting. The step function within R was utilized to make improvements to the model, which calculates the smallest AIC by identifying the optimal independent variables to be removed from the regression model. Due to the high number of variables, the AIC was used to critique the regression model relative to other models. The smaller the AIC value relative to other models, the better.

In the final step, the results of the new logistic regression model were evaluated that were generated from the step function. Firstly, the summary function was run against the new regression and it was noted that the number of independent variables was reduced to 23, including the dummy variables. This is a significant reduction in independent (predictor) variables compared to the original model. Next, three keyways to critique a logistic regression model were used: 1) p-values, 2) AIC, and 3) VIFs. The AIC is now 11,101, which is an improvement compared to the original model. Most of the remaining variables are identified as significant or reasonably close to the .05 p-value threshold. Please see summary output of the model below to see actual results.

```
glm(formula = readmit30 ~ sex + marital_status_c + insurance_provider +
    icu_yn + tobacco_user + drugabuse + mooddisorder + diabetes +
    anxiety + dementia + bmi + bp_diastolic + weight + height +
    pulse + ed_visits + ip_visits + chronic_conditions + los,
    family = binomial, data = hospital)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6060  -0.6680  -0.5670  -0.4494   2.3282

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                -5.912506   1.367870  -4.322 1.54e-05 ***
sexM                        0.206194   0.068416   3.014 0.002580 **
marital_status_csingle      0.151785   0.050642   2.997 0.002724 **
insurance_providermedicaid  0.124442   0.125011   0.995 0.319521
insurance_providermedicare  0.369094   0.080960   4.559 5.14e-06 ***
icu_yn1                     0.209894   0.108453   1.935 0.052948 .
tobacco_userNot Asked      -0.438991   0.402030  -1.092 0.274861
tobacco_userPassive        -0.344327   0.750900  -0.459 0.646555
tobacco_userQuit            0.203367   0.056367   3.608 0.000309 ***
tobacco_userYes             0.124297   0.083251   1.493 0.135425
drugabuse1                 -0.265008   0.133083  -1.991 0.046448 *
mooddisorder1               0.395063   0.194291   2.033 0.042016 *
diabetes1                   0.333704   0.181338   1.840 0.065734 .
anxiety1                    0.097677   0.058937   1.657 0.097456 .
dementia1                   0.204613   0.089336   2.290 0.022000 *
bmi                         0.060523   0.021116   2.866 0.004155 **
bp_diastolic               -0.005553   0.002102  -2.642 0.008248 **
weight                     -0.009315   0.003337  -2.791 0.005249 **
height                      0.042969   0.020471   2.099 0.035815 *
pulse                       0.007004   0.001623   4.314 1.60e-05 ***
ed_visits                   0.139154   0.009995  13.922  < 2e-16 ***
ip_visits                   0.178871   0.056568   3.162 0.001566 **
chronic_conditions          0.024008   0.004895   4.905 9.36e-07 ***
los                         0.053607   0.008034   6.673 2.51e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 11565  on 11862  degrees of freedom
Residual deviance: 11053  on 11839  degrees of freedom
AIC: 11101
```

Figure 1: Final model output

And finally, the VIFs were also evaluated – see figure 2 below for full results. For majority of the variables, the VIF values are below 5.0. There are just 3 variables that have high VIFs.

```
##       mooddisorder           diabetes              los
##           1.005024           1.013744         1.019922
##           dementia          ip_visits           icu_yn
##           1.020929           1.026711         1.026741
##          drugabuse              pulse     bp_diastolic
##           1.045073           1.059745         1.079153
##            anxiety          ed_visits marital_status_c
##           1.081965           1.087668         1.107680
## chronic_conditions insurance_provider      tobacco_user
##           1.119595           1.163424         1.178434
##                sex             height              bmi
##           2.043275          13.464283        56.210448
##             weight
##          63.560163
```

Figure 2: VIFs for each variable

Based on the lower AIC value most of the remaining variables had significant p- values and good VIF values, this model is considered a significant improvement compared to the original model. The improved regression was utilized to generate probabilities on another hospital data set called "hospital_new". This data set has 1,117 rows (each row represents a patient) and shares the same independent variables as the hospital data with an additional calculated probability based on our model. The calculated probability values in this new data set will help the user identify specific patients that are likely to be readmitted into a hospital within 30 days.

SECTION 3 - PATIENT DESCRIPTIONS

The profiles of the top 3 patients identified as most likely to be readmitted to the hospital within 30 days shared many of the same key indicators and a likelihood for readmittance of over 63%. The top 2 patients each had 13 visits to the emergency department (ED) within the past year, while the 3rd had 10 visits to the ED. This is at least 6 times higher than the average in the study. Our model found that, all other factors being equal, the odds of being readmitted within 30 days increased by 14% for each ED visit within the past 12 months.

Each of the top 3 patients also had a length of stay ranging between 6 and 10 days during their most recent hospitalization while the average was just 4.5 days. Each patient had also been diagnosed with between 14 to 16 chronic conditions, which is more than double the average of 6.6 chronic conditions. All other factors being equal, our model found that the odds of being readmitted within 30 days increased by 2% for every chronic condition noted. Finally, all 3 were identified as overweight or obese based on their body mass index (BMI), were former tobacco users, and were diagnosed with dementia or anxiety.

Like the top 3 patient profiles identified as most likely to be readmitted within 30 days of leaving the hospital, the bottom 3 found least likely to be readmitted shared several of the same key indicators. These include the type of health insurance, history of tobacco use, visits to the ED, and the number of diagnosed chronic conditions.

The bottom 3 patients identified as least likely to be readmitted within 30 days all had commercial health insurance. Our model found that, all other factors being equal, the odds of being readmitted within 30 days for a patient with Medicaid increased by 13%, and odds increased by 44% for those patients with Medicare. Additionally, the bottom 3 patients had just 4, 3, and 3 chronic conditions, respectively. Again, this is lower than the average of 6.6 chronic conditions in the study.

Length of hospital stay, and previous ED visits were also significant indicators for readmission. The bottom three patients had a length of stay of only 1 or 2 days in the hospital, while the average from the study was 4.5 days. Only 1 of the bottom 3 had been to the ED within the past year. Overall, the likelihood of these patients of being readmitted within 30 days is under 7%.