Wine sales prediction

Sri Seshadri

11/11/2017

Contents

1. Introduction

A large wine manufacturer is studying data collected on 12,000 commercially available wines, with a goal to predict the number of cases ordered based upon the characteristics. The manufacturer intends to adjust the winde offerings based on the findings. The data collected is related to chemical properties of wine with response variable being the number of sample cases sold to distribution companies. the data also includes features like review stars provided by the tasters and label appeal.

2. Exploratory Data analysis

The below table shows the summary statistics of the data. It is seen that there are missing data. "Stars" has the most missing values. The missing values mostly correspond to NO (0) cases sold. Which is most likely due to lack of opportunity to sample because of none sold.

Table 1: Summary Stats and missing values

	min	Q1	median	Q3	max	mean	sd	n	missing
INDEX	1.00	4037.50	8110.00	12106.50	16129.00	8069.98	4656.91	12795	0
TARGET	0.00	2.00	3.00	4.00	8.00	3.03	1.93	12795	0
FixedAcidity	-18.10	5.20	6.90	9.50	34.40	7.08	6.32	12795	0
VolatileAcidity	-2.79	0.13	0.28	0.64	3.68	0.32	0.78	12795	0
CitricAcid	-3.24	0.03	0.31	0.58	3.86	0.31	0.86	12795	0
ResidualSugar	-127.80	-2.00	3.90	15.90	141.15	5.42	33.75	12179	616
Chlorides	-1.17	-0.03	0.05	0.15	1.35	0.05	0.32	12157	638
FreeSulfurDioxide	-555.00	0.00	30.00	70.00	623.00	30.85	148.71	12148	647
TotalSulfurDioxide	-823.00	27.00	123.00	208.00	1057.00	120.71	231.91	12113	682
Density	0.89	0.99	0.99	1.00	1.10	0.99	0.03	12795	0
pH	0.48	2.96	3.20	3.47	6.13	3.21	0.68	12400	395
Sulphates	-3.13	0.28	0.50	0.86	4.24	0.53	0.93	11585	1210
Alcohol	-4.70	9.00	10.40	12.40	26.50	10.49	3.73	12142	653
LabelAppeal	-2.00	-1.00	0.00	1.00	2.00	-0.01	0.89	12795	0
AcidIndex	4.00	7.00	8.00	8.00	17.00	7.77	1.32	12795	0
STARS	1.00	1.00	2.00	3.00	4.00	2.04	0.90	9436	3359

Figure 2 shows the historgram of features in the data. The chemical properties of wines appear to share an identical distribution with peaks closer to zero. This mau be likely to some standardization done to the day. The variable "TARGET" looks to be poisson or negative biniomially distributed with inflation at 0.
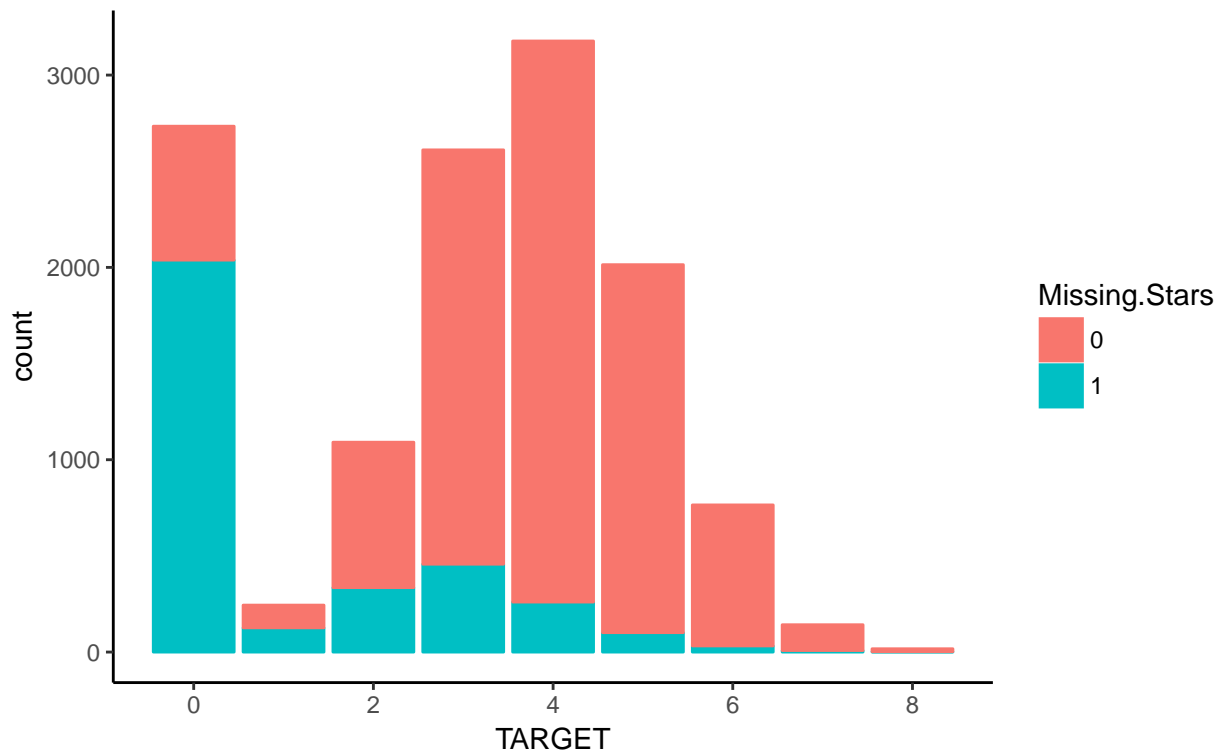
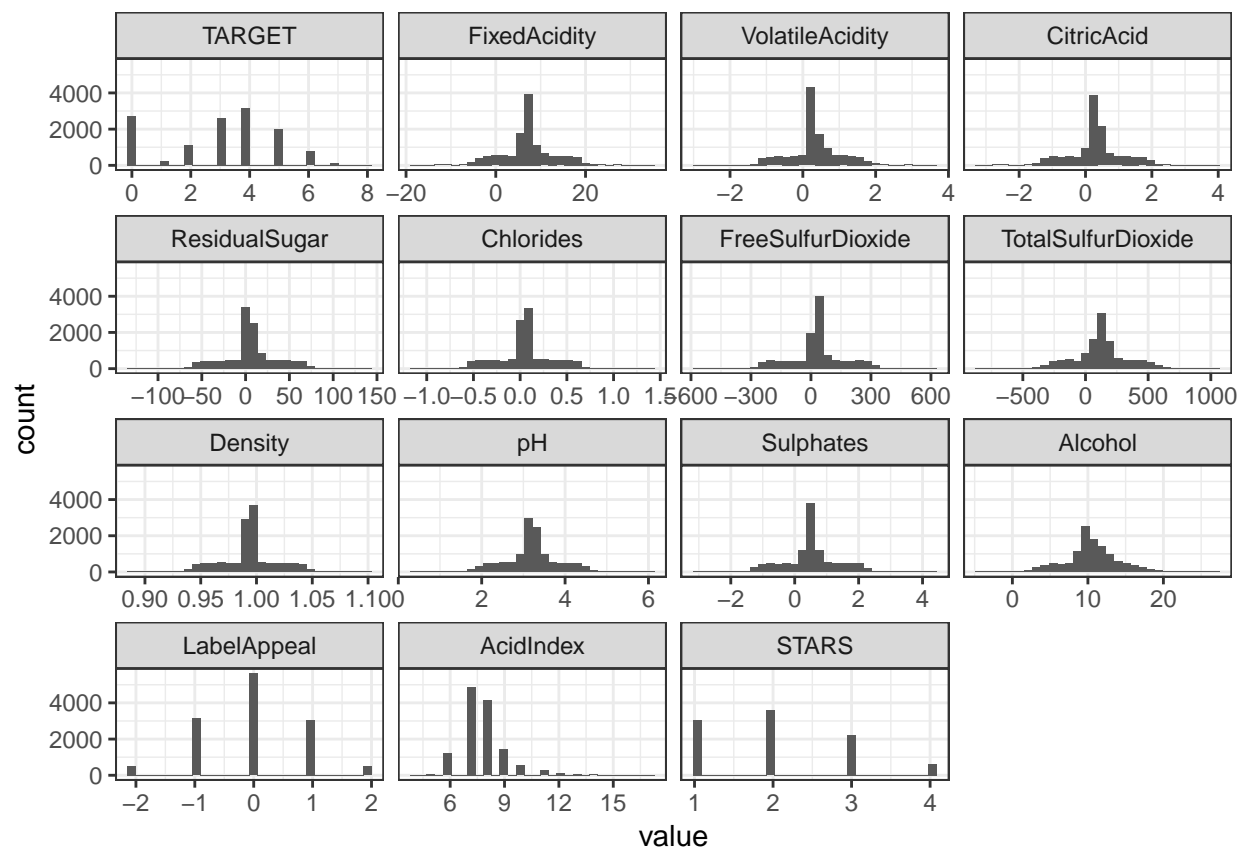Figure 1: Missing STARS values' association with Number of cases sold

Figure 2: Histograms of features

3

3. Feature selection

In this section we'll attempt to select important features that explain the target variable. We'll explore the correlations that might exist in the data. There is positive correlations between TARGET and STARS and LabelAppeal.
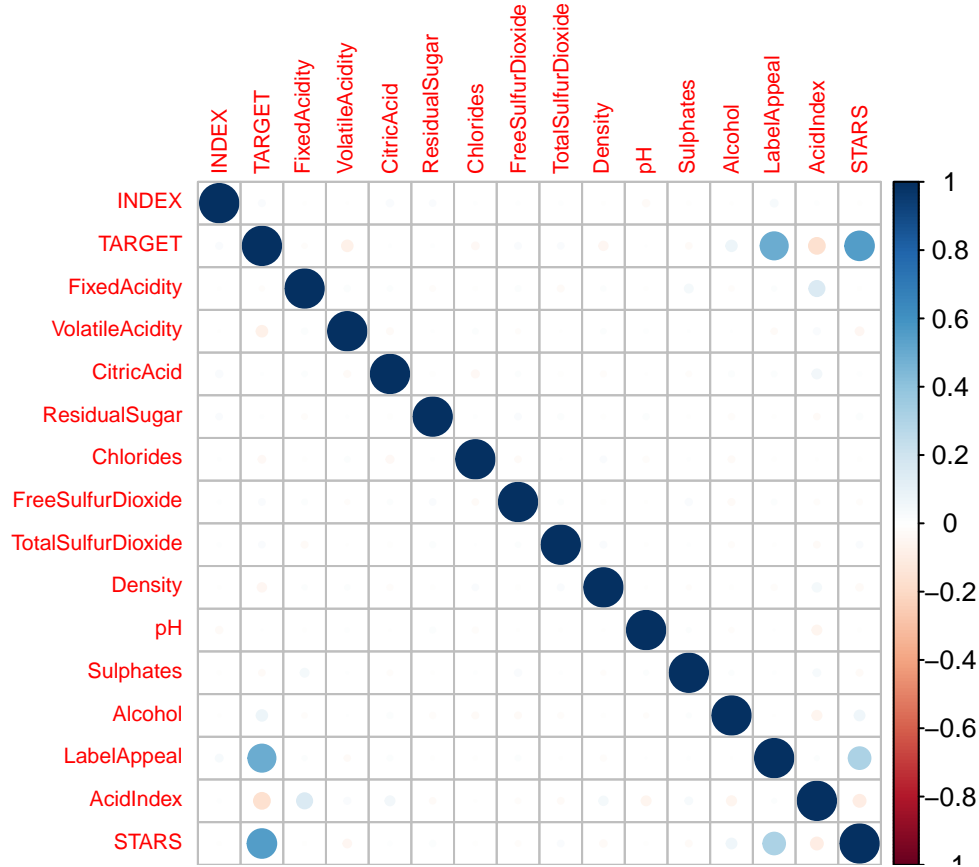

Figure 3: Correlation plot

It'll be useful to identify what contributes to the zero inflation in the TARGET variable. For which let's create an indicator variable "TARGET0" to equal 1 when TARGET is 0 and 0 otherwise.

3.1 Decision Trees

3.1.1 Predictors for zero cases sold (No sale)

Figure 4 shows the decision tree for TARGET0. Where 1 is no sale (cases sold = 0) and 0 is sale (cases sold > 0). Figure 5 shows the variable importance plot after a random forest bootstrap. While LabelAppeal did not contribute to node purity (Decrease in Gini index), it did affect accuracy on out of bag (OOB) samples.

3.1.2 Predictors for cases sold; when successfuly sold (cases > 0)

Figure 6 shows the decision tree of cases sold when they are greater than 0. It can be seen that the LabelAppeal and STARS are the top hitters. Figure 7 shows the variable importance plot when a random forest method is employed, where a random set of predictors are chosen at each iteration to fit a decision tree. Variables
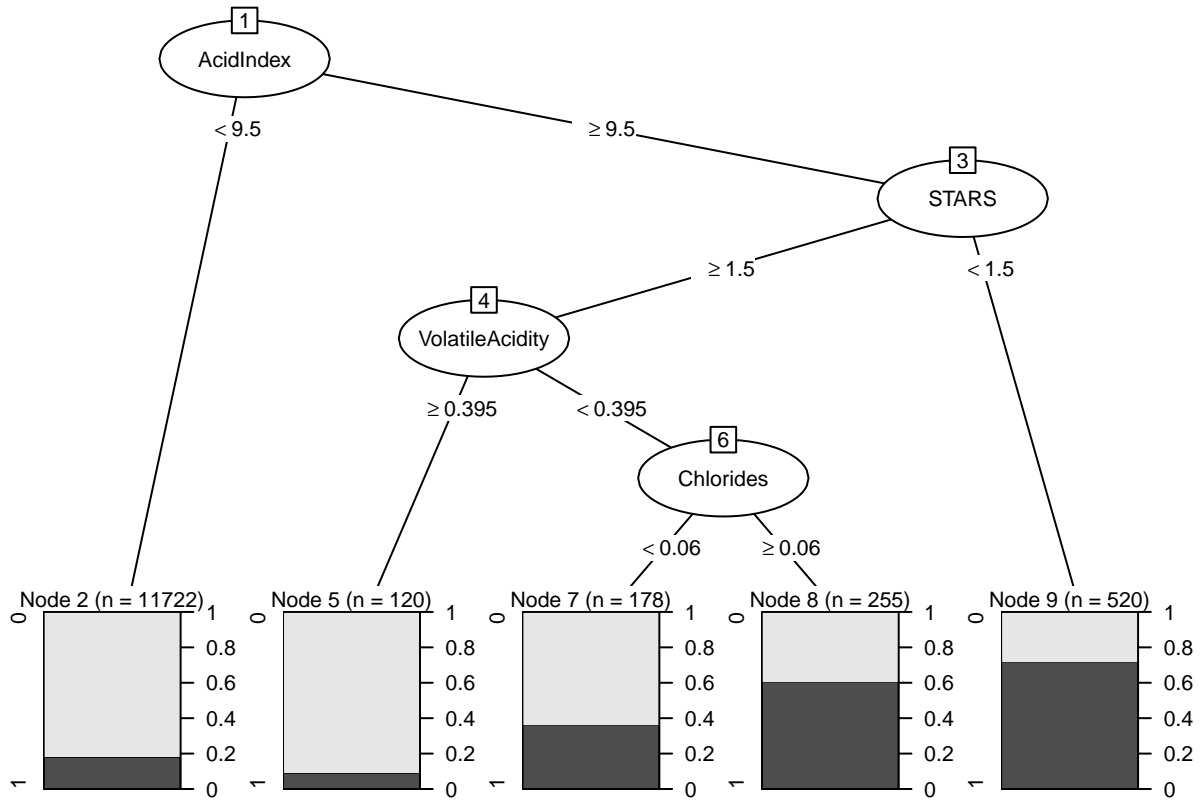
Figure 4: Decision tree - Sale (0) or no Sale (1)

whose exclusion contributes to higher Mean Squared Error (MSE) is deemed important. LabelApeal, STARS and Alcohol are top 3 variables that are important.
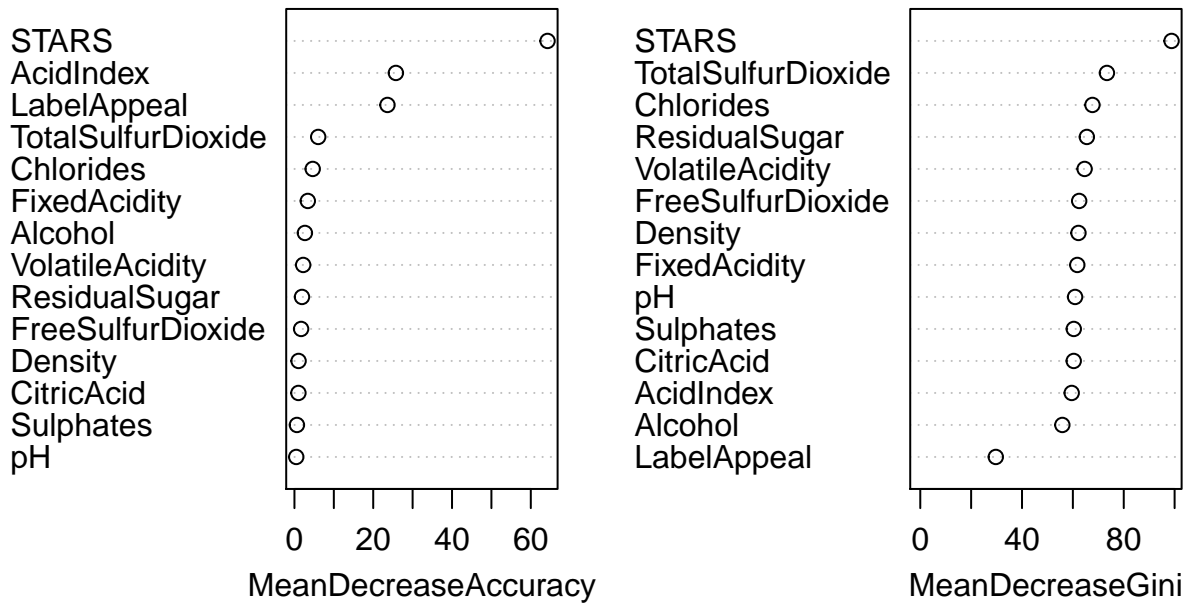
Variable Importance plot – Random forest
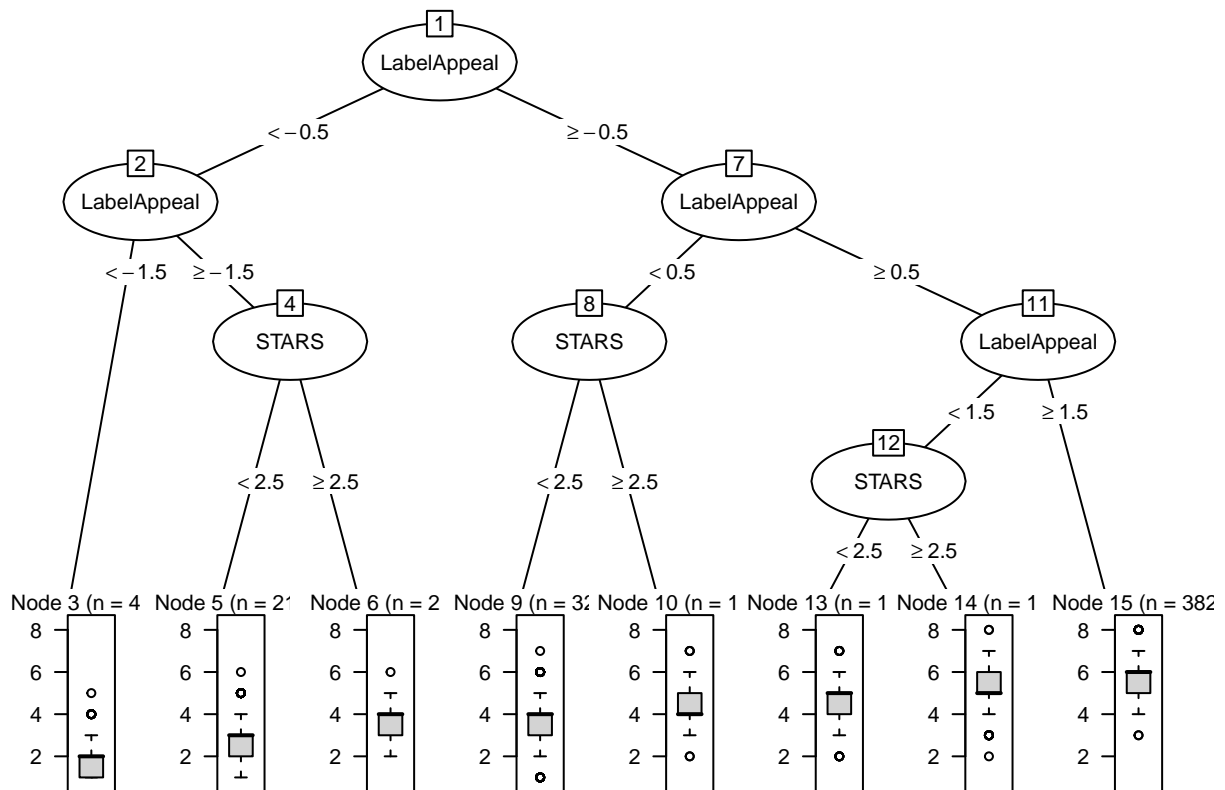

Figure 5: Variable Importance - Randomforest

Figure 6: Simple tree for TARGET

Variable importance plot from random forest – TARGET

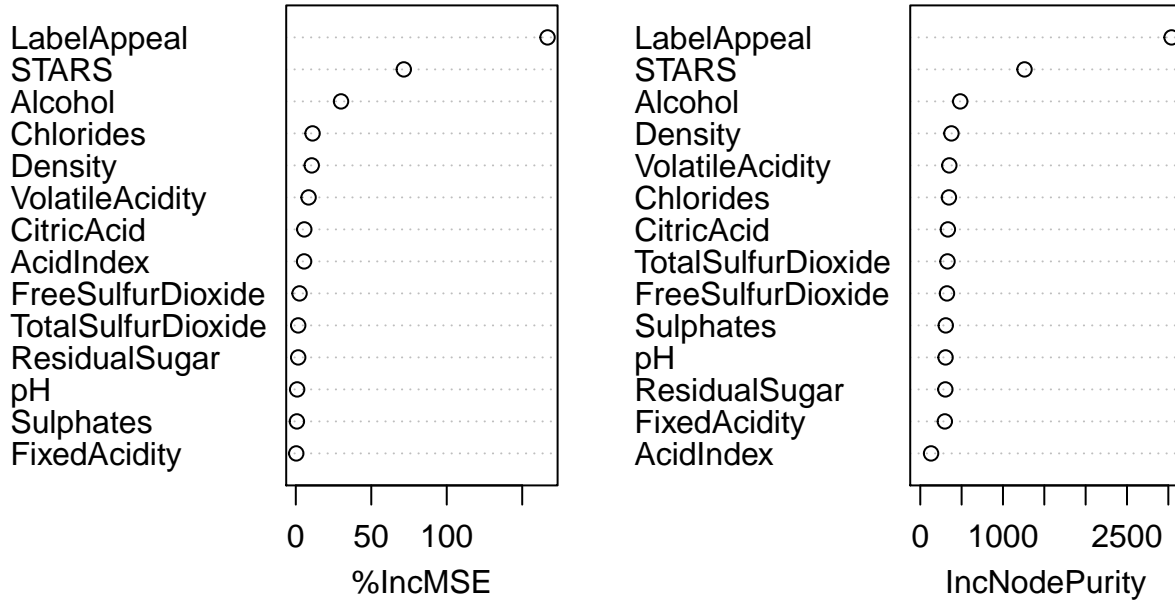|		
LabelAppeal		LabelAppeal
STARS		STARS
Alcohol		Alcohol
Chlorides		Density
Density		VolatileAcidity
VolatileAcidity		Chlorides
CitricAcid		CitricAcid
AcidIndex		TotalSulfurDioxide
FreeSulfurDioxide		FreeSulfurDioxide
TotalSulfurDioxide		Sulphates
ResidualSugar		pH
pH		ResidualSugar
Sulphates		FixedAcidity
FixedAcidity		AcidIndex

%IncMSE   IncNodePurity

Figure 7: Variable importance plot - TARGET

4. Training and validation samples

The data is split into training and validation sample for evaluating models for final selection. The 75% of the data is sampled as training sample and the rest is set aside as validation set. Figure 8 shows that the TARGET variable is identically distributed for both training and validation samples.

5. Modeling

The following modeling approaches will be tried for predicting the number of cases sold.

- Poisson regression
- Negative binomial regression
- Zero infalted Poisson regression
- Zero inflated Negative binomial regression
- OLS regression

5.1 Poisson regression.

As seen in table 1, the mean and standard deviation of TARGET are not too far apart. Poisson model is a good candidate for regression modeling.
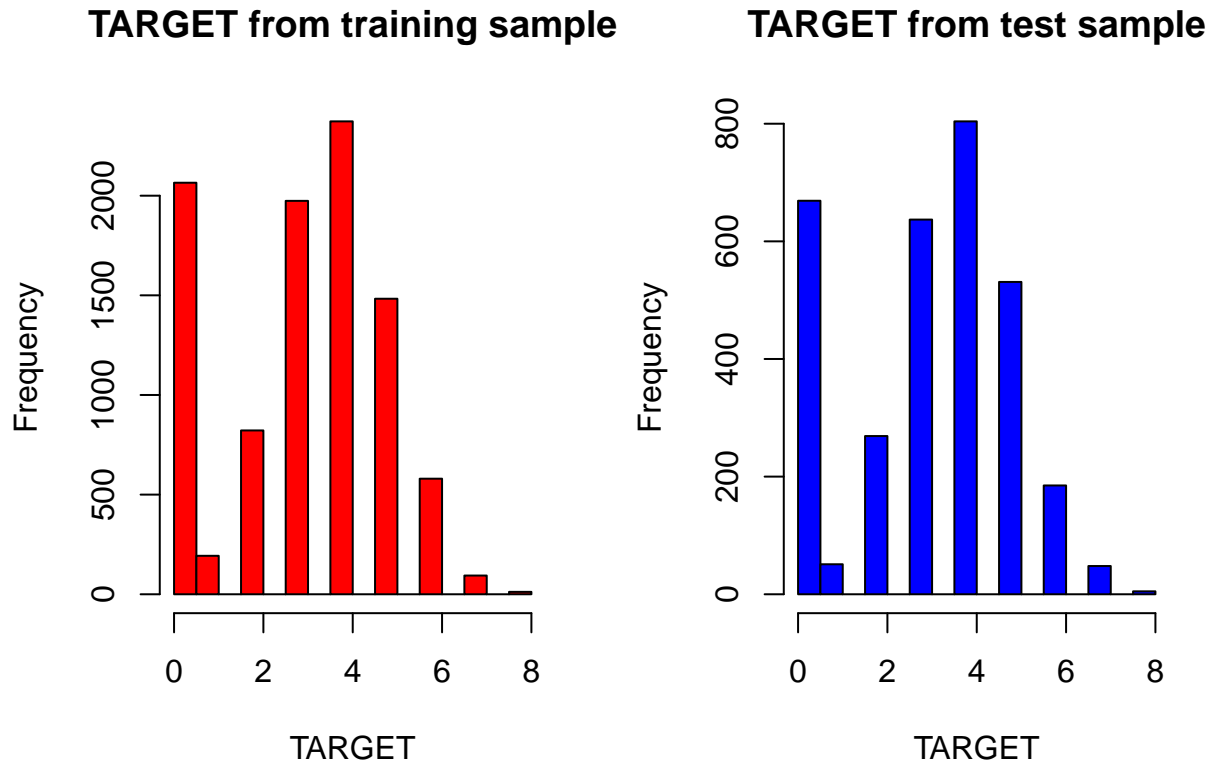
Figure 8: TARGET distribution for training and validation samples

5.1.1 Simple model with Label appeal as the predictor.

While we know that the TARGET variable is zero inflated and there are atleast more than one important predictor from the section above, we'll attempt to build the model ground up with LabelAppeal as the single predictor.

```
##
## Call:
## glm(formula = TARGET ~ LabelAppeal, family = poisson(link = "log"),
##     data = train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.1342  -0.4281   0.1717   0.5821   2.0412
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.080397   0.006012  179.70   <2e-16 ***
## LabelAppeal 0.255594   0.006623   38.59   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 17183  on 9595  degrees of freedom
## Residual deviance: 15689  on 9594  degrees of freedom
## AIC: 39571
```

```
## 
## Number of Fisher Scoring iterations: 5

## Likelihood ratio test
## 
## Model 1: TARGET ~ LabelAppeal
## Model 2: TARGET ~ 1
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1   2 -19783
## 2   1 -20530 -1 1494.4  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 2: Simple Poisson model statistics

null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual	pseudoR.Sq	MAE
17182.9	9595	-19783.35	39570.7	39585.04	15688.51	9594	0.09	1.39

5.1.1.1 Model interpretation

It can be seen from the Likelihood ratio test ab0ve, that the slope of LabelAppeal is NOT zero and an unit increase in label appeal increases by 1. However the model is not an adequate fit with residual deviance over degrees of freedom being much higher than 1 (in comparison to a saturated model). This is also reflected in the Pseudo R Squared value.

The model does not predict 0s. As can be see from figure 9.

5.1.2 Poisson model with Label appeal & Stars as predictors

Now, we'll use Label appeal and stars as predictors and compare the model with the simple model in the above section.

```
## 
## Call:
## glm(formula = TARGET ~ LabelAppeal + STARS, family = poisson(link = "log"),
##     data = train[!is.na(train$LabelAppeal), ])
## 
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.16162  -0.26871   0.04889   0.36797   1.65061
## 
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.869060   0.016581    52.41   <2e-16 ***
## LabelAppeal 0.179589   0.007595    23.65   <2e-16 ***
## STARS       0.190392   0.007074    26.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for poisson family taken to be 1)
## 
##     Null deviance: 6398.3  on 7057  degrees of freedom
## Residual deviance: 4457.8  on 7055  degrees of freedom
##   (2538 observations deleted due to missingness)
```
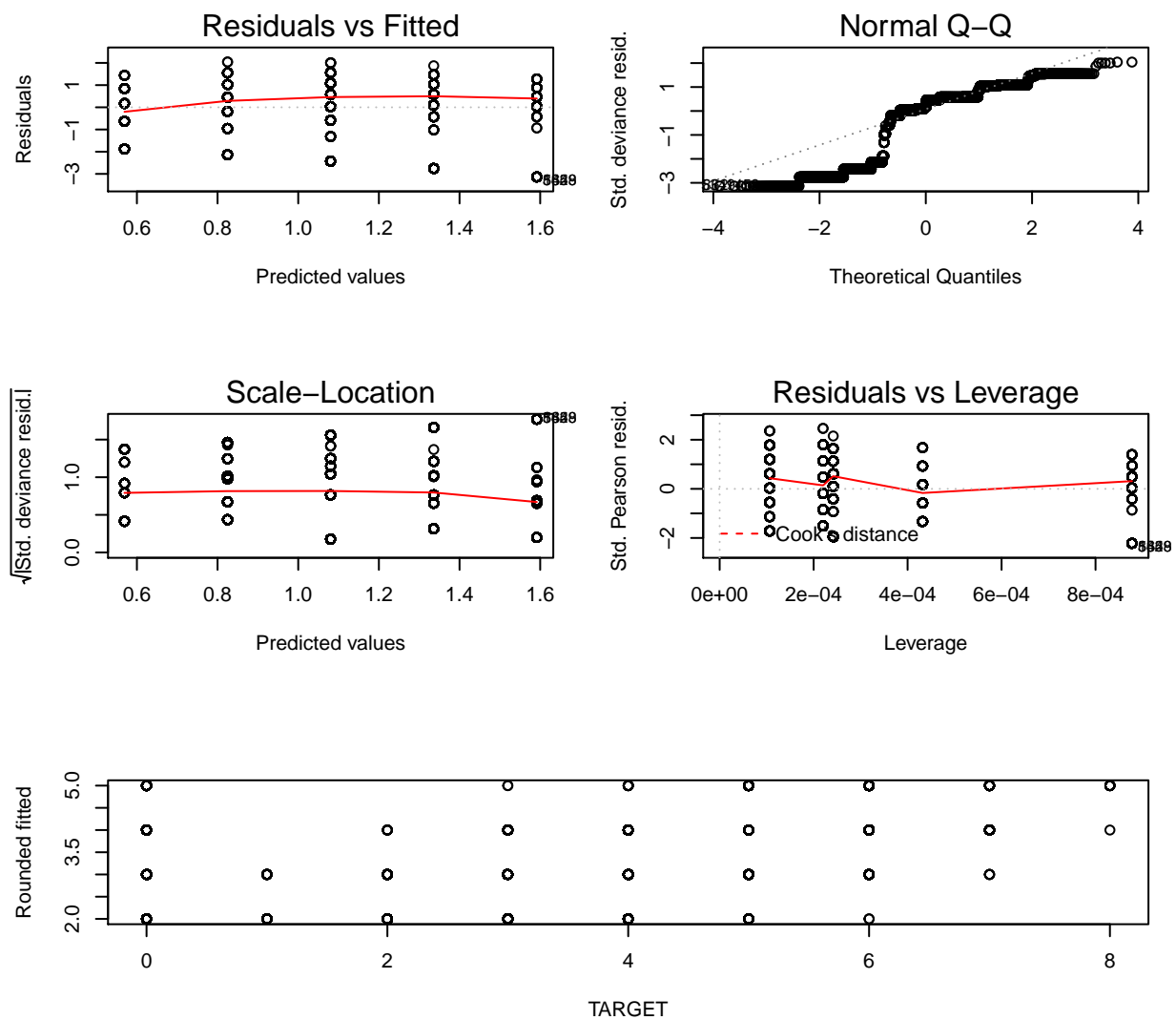
Figure 9: Simple poisson regression diagnostics

```
## AIC: 25458
##
## Number of Fisher Scoring iterations: 4

## Likelihood ratio test
##
## Model 1: TARGET ~ LabelAppeal + STARS
## Model 2: TARGET ~ 1
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1   3 -12726
## 2   1 -13696 -2 1940.5  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 3: Poisson model statistics with STARS & LabelAppeal as predictors

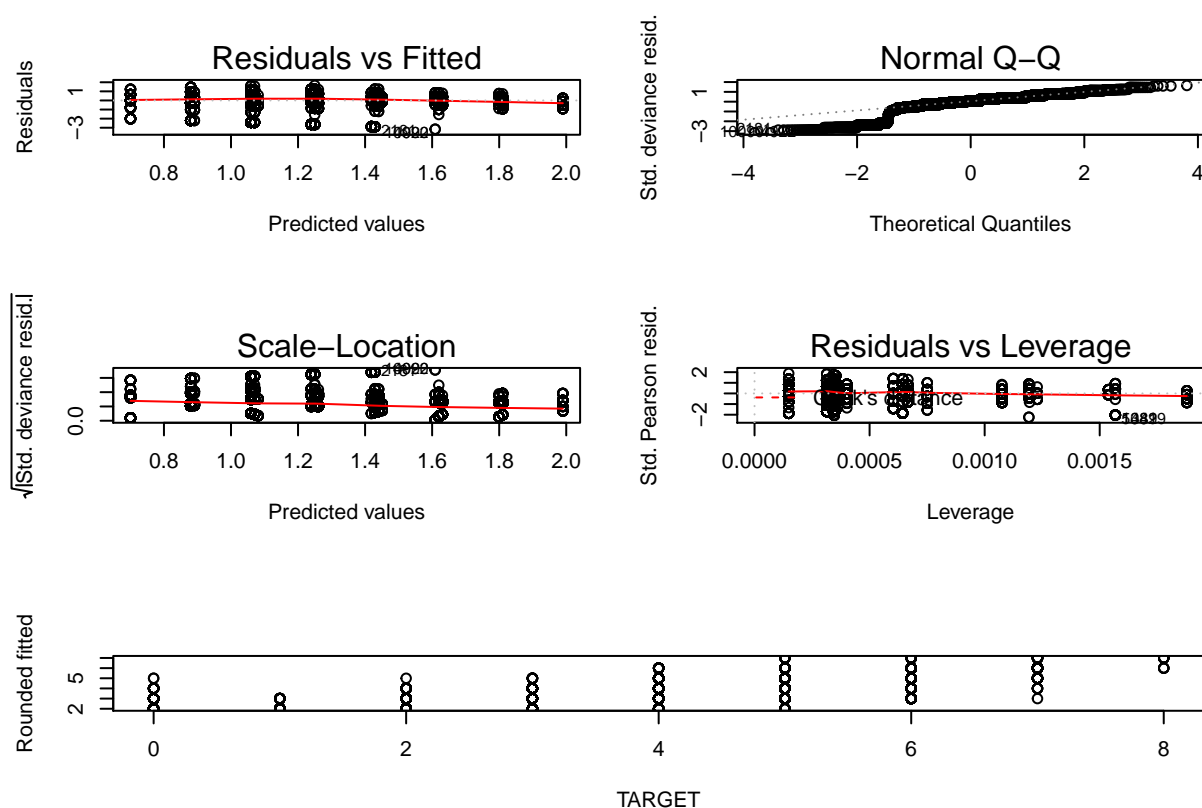| null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual	pseudoR.Sq	MAE
6398.3	7057	-12726.13	25458.26	25478.85	4457.75	7055	0.3	0.88

Figure 10: Model diagnostics TARGET ~ LabelAppeal + STARS

5.1.2.1 Model interpretation.

It is seen that the regression model is statistically significant from the Likelihood Ratio Test (LRT) and the regression coefficents are also significant. The MAE is 1.5 which is a marginally higher than the simple

model with LabelAppeal as predictor. However the ratio of residual deviance to its degrees of freedom is significantly less than 1, which can be interpreted as better model fit to the data. This model also does not fit the zero counts.

5.1.3 Dropping one predcitor from model.

It'll be interesting to see if there is value in dropping one of the regressors from the model. Below, LRT is performed by dropping one regressor at a time from the model.

```
## Single term deletions
##
## Model:
## TARGET ~ LabelAppeal + STARS
##             Df Deviance   AIC    LRT  Pr(>Chi)
## <none>          4457.8 25458
## LabelAppeal  1   5017.2 26016 559.40 < 2.2e-16 ***
## STARS        1   5174.8 26173 717.07 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From LRT, both Label Appeal and STARS appear significant. Dropping Label Appeal (kepping STARS as the only predictor) providing a marginal increase in model fit compared to dropping STARS(keeping Label Appeal as the only predictor). This interpretation is onsistent with the decision tree interpretation in the section 3.1.2.

The LRT test for model comparison may be biased, since the models may have different samples sizes due to missing data. However, LRT can be used as a guidance for model selection. It is examined in the below section just to be sure

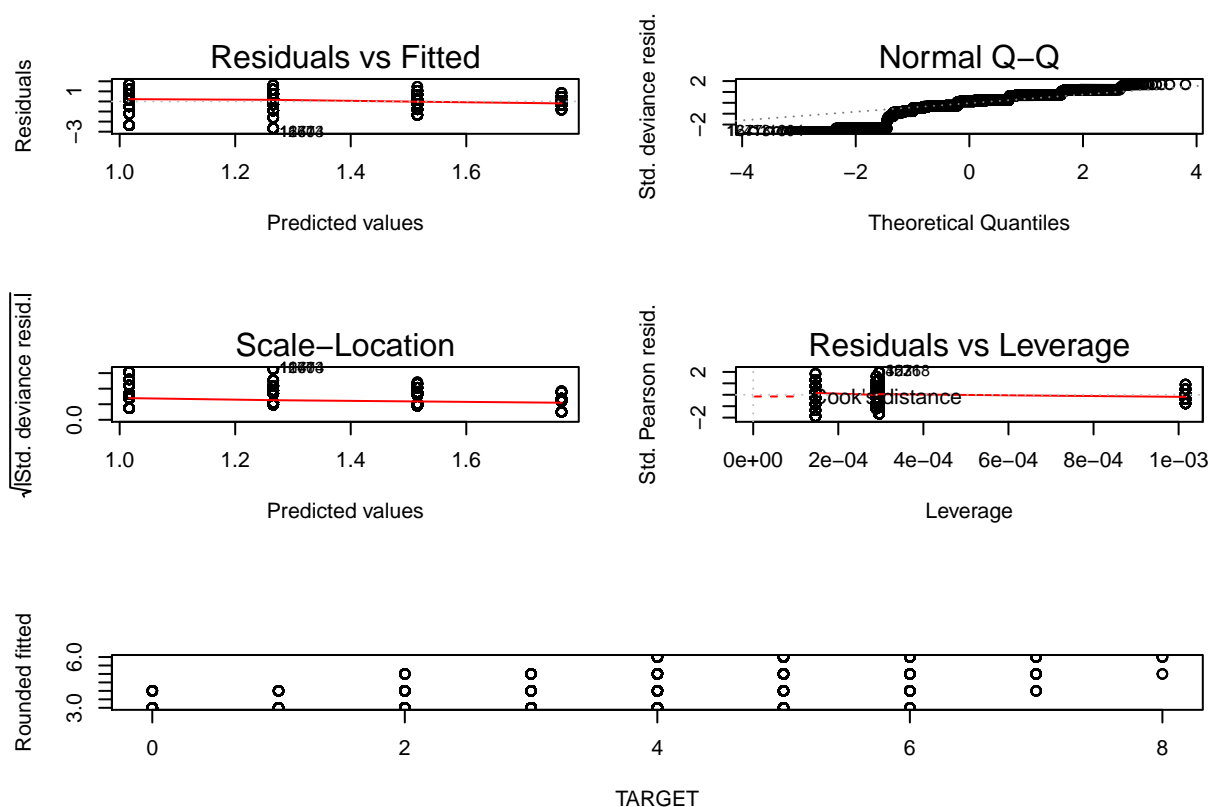**5.1.3.1 Two models - A. STARS as predictor B. LabelAppeal and Interpretation.**

Verifying the results in the above section to make sure the drop1 test is providing us the expected results. It can be seen that when the data is kept contant between the models, the results are as expected. The result for model B is different from that of in section 5.1.1, this is because the data used to fit the model is different. Based on AIC and MAE comparison Label Appeal seems to be a better predcitor so far, when using a Poisson model.

```
##
## Call:
## glm(formula = TARGET ~ STARS, family = poisson(link = "log"),
##     data = train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6632  -0.2980   0.1404   0.2361   1.6824
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.766962   0.016161   47.46   <2e-16 ***
## STARS       0.249465   0.006639   37.58   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
```

```
##     Null deviance: 6398.3  on 7057  degrees of freedom
## Residual deviance: 5017.2  on 7056  degrees of freedom
##   (2538 observations deleted due to missingness)
## AIC: 26016
##
## Number of Fisher Scoring iterations: 5

## Likelihood ratio test
##
## Model 1: TARGET ~ STARS
## Model 2: TARGET ~ 1
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1   2 -13006
## 2   1 -13696 -1 1381.1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 4: Simple Poisson model statistics with STARTS as regressor

null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual	pseudoR.Sq	MAE
6398.3	7057	-13005.83	26015.66	26029.39	5017.16	7056	0.22	0.97

Figure 11: Model diagnostics TARGET ~ STARS

```
##
## Call:
## glm(formula = TARGET ~ LabelAppeal, family = poisson(link = "log"),
```

13

```
##     data = train[complete.cases(train[, c(14, 16)]), ])
##
## Deviance Residuals:
##     Min      1Q   Median      3Q     Max
## -3.4169  -0.3004   0.2063   0.2336   1.6783
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.267168   0.006425  197.24   <2e-16 ***
## LabelAppeal 0.248567   0.007119   34.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 6398.3  on 7057  degrees of freedom
## Residual deviance: 5174.8  on 7056  degrees of freedom
## AIC: 26173
##
## Number of Fisher Scoring iterations: 4

## Likelihood ratio test
##
## Model 1: TARGET ~ LabelAppeal
## Model 2: TARGET ~ 1
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1   2 -13085
## 2   1 -13696 -1 1223.5  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 5: Simple Poisson model statistics with LabelAppeal as regressor; non missing cases for STARS and LabelAppeal

null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual	pseudoR.Sq	MAE
6398.3	7057	-13084.67	26173.33	26187.06	5174.83	7056	0.19	0.85

5.1.4 Poisson regression with LabelAppeal, STARS and Alcohol as predictors

In this section Alcohol is added as a preditor to the model in 5.1.2. Then LRT is performed by dropping one variable from the predictor. Though Alcohol is statistically significant, it does not add significant predictive power to the model. There is no significant reduction in MAE. Also we see that there is pattern in residuals based on the levels of the TARGET variable. It is to be noted that the predicted values are in log scale.

```
##
## Call:
## glm(formula = TARGET ~ LabelAppeal + STARS + Alcohol, family = poisson(link = "log"),
##     data = train[complete.cases(train[, c(13, 14, 16)]), ])
##
## Deviance Residuals:
##     Min      1Q   Median      3Q     Max
## -3.2162  -0.2634   0.0512   0.3702   1.6226
```
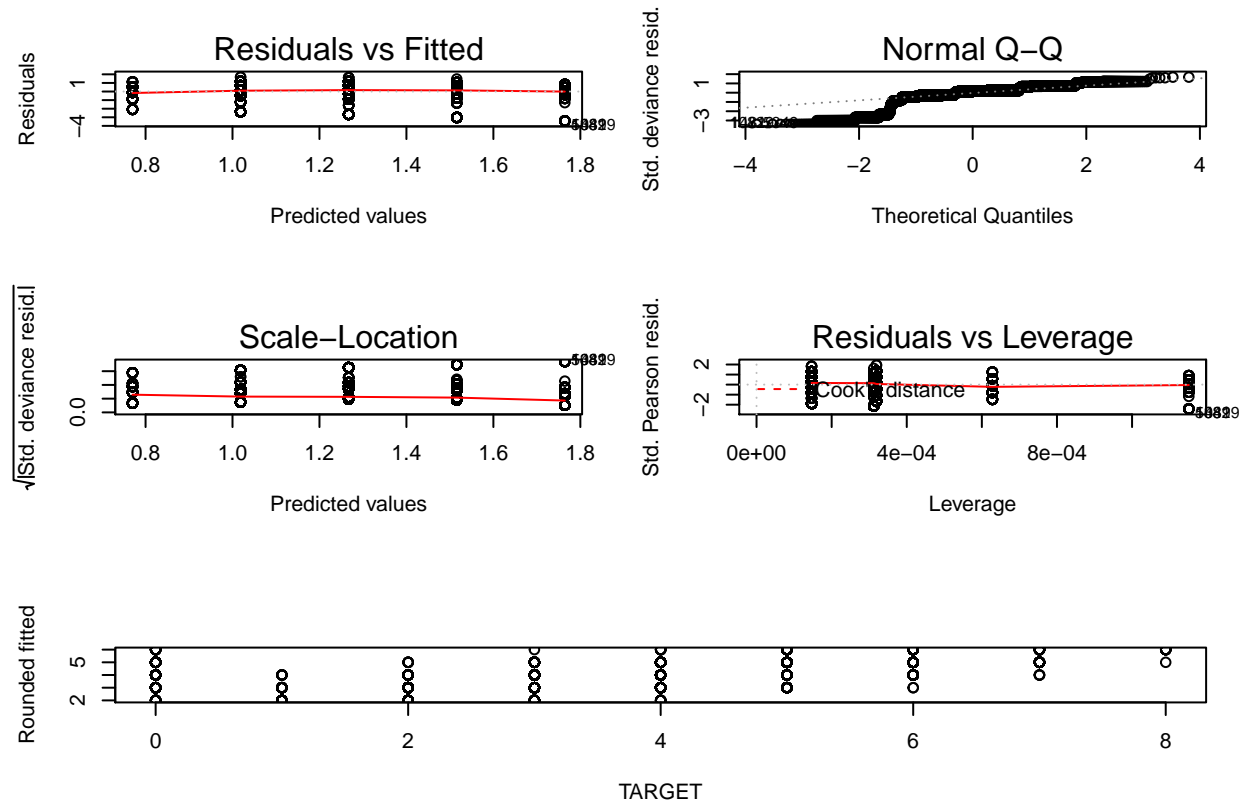
Figure 12: Model diagnostics TARGET ~ LABEL APPEAL non missing data for STARS and LabelAppeal

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.812942   0.024086  33.751   <2e-16 ***
## LabelAppeal 0.179440   0.007813  22.967   <2e-16 ***
## STARS       0.189742   0.007316  25.934   <2e-16 ***
## Alcohol     0.005377   0.001722   3.122   0.0018 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 6076.9  on 6697  degrees of freedom
## Residual deviance: 4230.5  on 6694  degrees of freedom
## AIC: 24157
##
## Number of Fisher Scoring iterations: 4

## Likelihood ratio test
##
## Model 1: TARGET ~ LabelAppeal + STARS + Alcohol
## Model 2: TARGET ~ 1
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1   4 -12074
## 2   1 -12998 -3 1846.4  < 2.2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 6: Poisson model statistics with LabelAppeal, STARS, AL-
COHOL as regressors

null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual	pseudoR.Sq	MAE
6076.92	6697	-12074.49	24156.99	24184.22	4230.55	6694	0.3	0.83
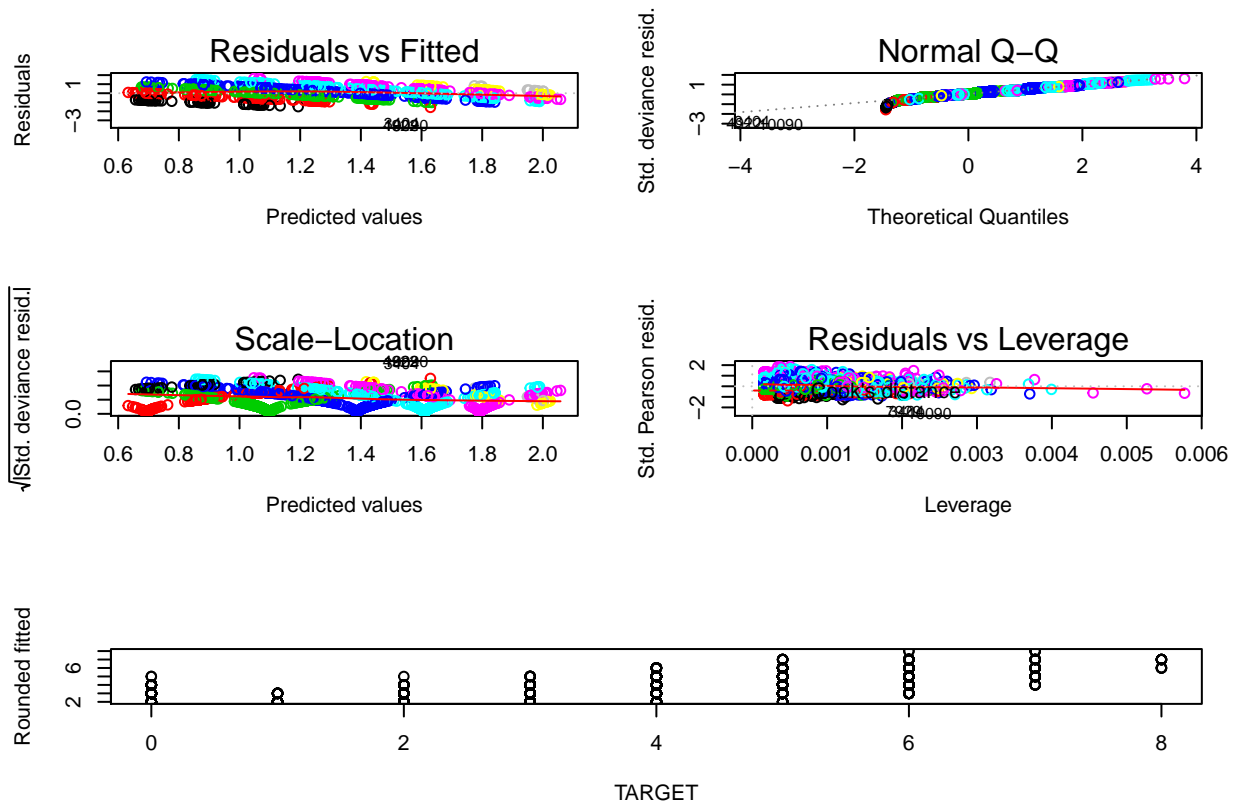
5.1.4.1 Model interpretation

```
## Single term deletions
##
## Model:
## TARGET ~ LabelAppeal + STARS + Alcohol
##               Df Deviance   AIC    LRT  Pr(>Chi)
## <none>           4230.5 24157
## LabelAppeal   1   4758.2 24683 527.70 < 2.2e-16 ***
## STARS         1   4896.6 24821 666.01 < 2.2e-16 ***
## Alcohol       1   4240.3 24165   9.75  0.001796 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From LRT by droping one predictor at every iteration, it is seen that the exclusion of Alcohol does not affect
the fit considerably. Hence it is best to drop Alcohol from the model.

5.1.5 Addition of AcidIndex as predictor

One of the other predictors that was found key in section 3.1.2 is AcidIndex. It is seen below that though Acidindex's coefficient is statsitically significant, it does not improve prediction. Alcohol's coefficient is marginally significant.

```
##
## Call:
## glm(formula = TARGET ~ LabelAppeal + STARS + Alcohol + AcidIndex,
##     family = poisson(link = "log"), data = train[complete.cases(train[,
##         c(13, 14, 15, 16)]), ])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2171  -0.2763   0.0566   0.3687   1.6568
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.200847   0.052053  23.070   <2e-16 ***
## LabelAppeal  0.181869   0.007820  23.257   <2e-16 ***
## STARS        0.183927   0.007352  25.017   <2e-16 ***
## Alcohol      0.004418   0.001725   2.560   0.0105 *
## AcidIndex   -0.048170   0.005763  -8.359   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 6076.9  on 6697  degrees of freedom
## Residual deviance: 4158.8  on 6693  degrees of freedom
## AIC: 24087
##
## Number of Fisher Scoring iterations: 5

## Likelihood ratio test
##
## Model 1: TARGET ~ LabelAppeal + STARS + Alcohol + AcidIndex
## Model 2: TARGET ~ 1
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1   5 -12039
## 2   1 -12998 -4 1918.2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 7: Poisson model statistics with LabelAppeal, STARS, AL-COHOL, AcidIndex as regressors

null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual	pseudoR.Sq	MAE
6076.92	6697	-12038.6	24087.2	24121.25	4158.77	6693	0.32	0.81

```
## Single term deletions
##
## Model:
## TARGET ~ LabelAppeal + STARS + Alcohol + AcidIndex
```

```
##                Df Deviance   AIC    LRT  Pr(>Chi)
## <none>              4158.8 24087
## LabelAppeal  1     4699.9 24626 541.13  < 2e-16 ***
## STARS        1     4778.7 24705 619.93  < 2e-16 ***
## Alcohol      1     4165.3 24092   6.56  0.01044 *
## AcidIndex    1     4230.5 24157  71.78  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

drop1 test shows that the exclusion of Alcohol provides the best AIC. In the next section we will fit a model with LabelAppeal, AcidIndex and STARS as predictors.

5.1.5.1 Removal of Alcohol and inclusion of AcidIndex

The model with LabelAppeal, STARS and AcidIndex is fit below and it is seen that all the predictor variables' coefficients are statistically significant. However the MAE is not improved. LRT when one predictor is left out of the model at each iteration shows that inclusion of AcidIndex does NOT impact the AIC considerably.

```
##
## Call:
## glm(formula = TARGET ~ LabelAppeal + STARS + AcidIndex, family = poisson(link = "log"),
##     data = train[complete.cases(train[, c(14, 15, 16)]), ])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2163  -0.2567   0.0445   0.3873   1.6245
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.251460   0.046855   26.71   <2e-16 ***
## LabelAppeal  0.182091   0.007602   23.95   <2e-16 ***
## STARS        0.184124   0.007113   25.89   <2e-16 ***
## AcidIndex   -0.048649   0.005605   -8.68   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 6398.3  on 7057  degrees of freedom
## Residual deviance: 4380.3  on 7054  degrees of freedom
## AIC: 25383
##
## Number of Fisher Scoring iterations: 5

## Likelihood ratio test
##
## Model 1: TARGET ~ LabelAppeal + STARS + AcidIndex
## Model 2: TARGET ~ 1
##   #Df LogLik Df Chisq Pr(>Chisq)
## 1   4 -12687
## 2   1 -13696 -3  2018  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 8: Poisson model statistics with LabelAppeal, STARS, AL-COHOL, AcidIndex as regressors

null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual	pseudoR.Sq	MAE
6398.3	7057	-12687.43	25382.86	25410.3	4380.35	7054	0.32	0.82

5.1.5.1 Model interpretation

Droping one predictor and performing the LRT, it can is seen below that removing AcidIndex does NOT move the needle on AIC.

```
## Single term deletions
##
## Model:
## TARGET ~ LabelAppeal + STARS + AcidIndex
##            Df Deviance   AIC    LRT  Pr(>Chi)
## <none>         4380.3 25383
## LabelAppeal  1   4954.3 25955 573.95 < 2.2e-16 ***
## STARS        1   5043.8 26044 663.49 < 2.2e-16 ***
## AcidIndex    1   4457.8 25458  77.41 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5.1.6 Poisson regression summary

```
##
## ------------------------------------------------------------------------------
##           model              null.deviance  df.null  logLik   AIC
## ------------------------------------------------------------------------------
##     TARGET = 1.08 + 0.256 *      17183        9595   -19783   39571
##         LabelAppeal
##
##     TARGET = 0.869 + 0.18 *       6398        7057   -12726   25458
##    LabelAppeal + 0.19 * STARS
##
##  TARGET = 0.767 + 0.249 * STARS   6398        7057   -13006   26016
##
##     TARGET = 1.267 + 0.249 *      6398        7057   -13085   26173
##         LabelAppeal
##
##     TARGET = 0.813 + 0.179 *      6077        6697   -12074   24157
##   LabelAppeal + 0.19 * STARS +
##       0.005 * Alcohol
##
##     TARGET = 1.201 + 0.182 *      6077        6697   -12039   24087
##   LabelAppeal + 0.184 * STARS +
##    0.004 * Alcohol - 0.048 *
##         AcidIndex
##
##     TARGET = 1.251 + 0.182 *      6398        7057   -12687   25383
##   LabelAppeal + 0.184 * STARS -
##       0.049 * AcidIndex
## ------------------------------------------------------------------------------
```

```
## 
## Table: Poisson fit summary (continued below)
## 
## 
## ---------------------------------------------------
##   BIC     deviance   df.residual   pseudoR.Sq   MAE
## -------  ----------  -------------  -----------  -----
## 39585     15689         9594          0.09        1
## 
## 25479      4458         7055          0.3         0.9
## 
## 26029      5017         7056          0.2         1
## 
## 26187      5175         7056          0.2         0.9
## 
## 24184      4231         6694          0.3         0.8
## 
## 24121      4159         6693          0.3         0.8
## 
## 25410      4380         7054          0.3         0.8
## ---------------------------------------------------
```