Solution to the Logistic Regression Graded Assignment

After Analyzing the data sets and our client requirement I came to know the following details.
Independent Variables:
X2 Brand A chips: Are made with farm grown ingredients like potato, corn or wheat?
X9 Brand A chips: Have zero grams trans fat
X16 Brand A chips: Are made with natural oils
X30 Brand A chips : 10=minimally Processed / 1=Heavily processed on a 10 point scale

Dependent variable:
X23 Brand A chips : Rate the following 10=good for you, 1=bad for you

As X23 has 10 levels from 1 to 10,I assumed greater than 5 are good and less than that are bad,and created another categorical variable called target with 0 and 1,where 0 is bad and 1 is good. After Data preparation and exploration,I came up with the following final model equation with the help of our Logistic Regression function.

Model Equation:
log(P/1-P)= -1.43972 +(-0.38741)*X2+(-0.34811)*X9 +(-0.43805)*X16+ 0.42569*X30
P=e^-1.43972 +(-0.38741)*X2+(-0.34811)*X9 +(-0.43805)*X16+ 0.42569*X30/(1+ e^-1.43972 +(-0.38741)*X2+(-0.34811)*X9 +(-0.43805)*X16+ 0.42569*X30)
Insights:
From the model we can infer that except from the processing level all other have -ve impact
Farm grown ingredients has a -ve impact of 0.38741 on the Brand value
Have zero grams trans fat has -ve impact of 0.34811 on the brand value
Are made with natural oils has -ve impact of 0.43805 on the brand value
Processing level has a positve impact of 0.42569 on the brand value

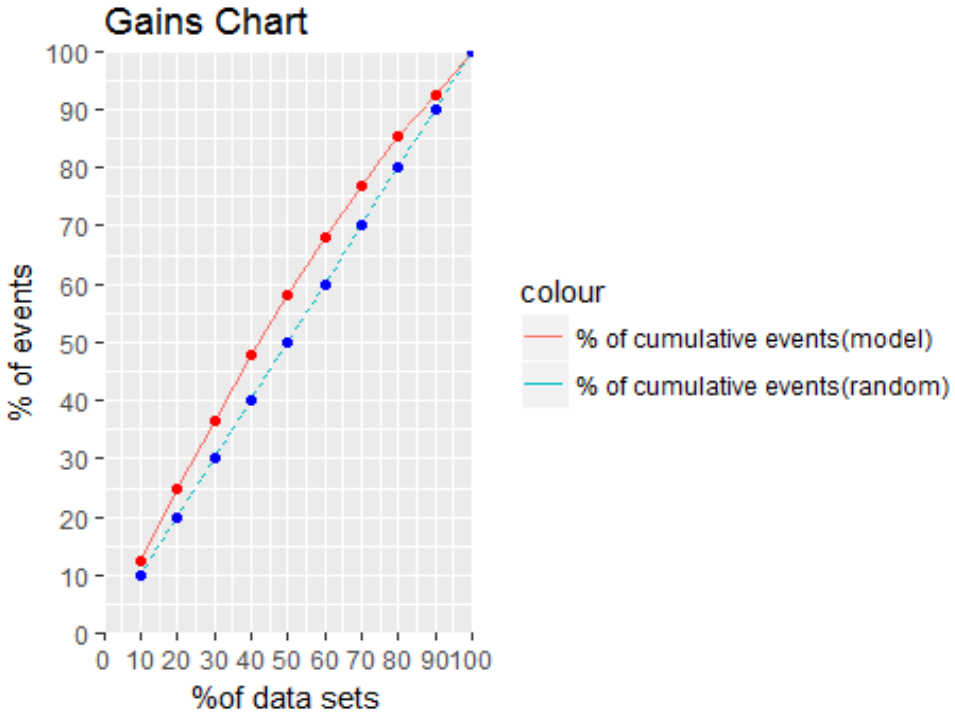
Kappa,Accuracy,Gains Chart,Lift Chart:
For validating the model I checked the accuracy and kappa value and optimized it using the confusion matrix and plotted Gains Chart and Lift Chart,details has been given below.

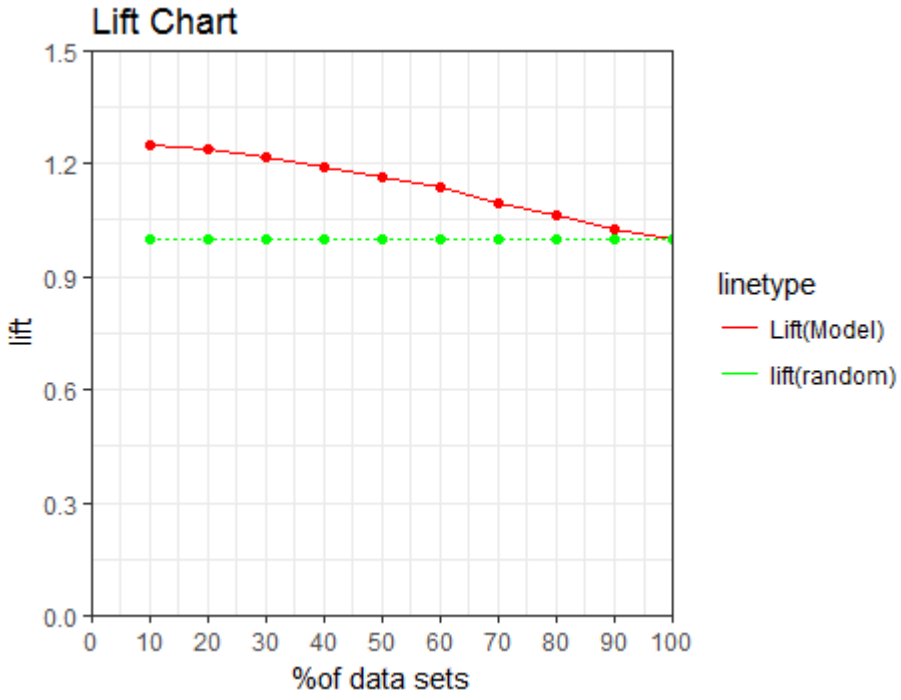
After optimizing the kappa value ,I went with its corresponding Cutoff value to predict which is my good and bad.And my cutoff value was coming to be 0.39 for which I got the maximum kappa and accuracy.So I went with
formula: {test$result<-ifelse(pred>0.39,1,0)
confusionMatrix(test$result,test$target,positive = "1")}
Accuracy is coming to be 76.89%,which seems to be fine.

```
Confusion Matrix and Statistics

    Reference
Prediction    0    1
    0 4694  988
    1  684  869

      Accuracy : 0.7689
        95% CI : (0.759, 0.7786)
  No Information Rate : 0.7433
  P-Value [Acc > NIR] : 2.577e-07
```

```
       Kappa : 0.3601
Mcnemar's Test P-Value : 1.262e-13

     Sensitivity : 0.4680
     Specificity : 0.8728
   Pos Pred Value : 0.5596
   Neg Pred Value : 0.8261
      Prevalence : 0.2567
   Detection Rate : 0.1201
Detection Prevalence : 0.2147
  Balanced Accuracy : 0.6704

   'Positive' Class : 1
```

My Gains Chart and Lift chart shows that my model is better than the random model




Lift Chart
1.5
1.2
0.9
0.6
0.3
0.0
lift
0
10
20
30
40
50
60
70
80
90
100
%of data sets
linetype
Lift(Model)
lift(random)