Logistic Regression Modeling (Cleaned_3.csv)

Ziyan Zhao

02/09/2020

Attributes:

1: Id
2: StageName (factor)
3: Status_Reason___c
4: RecordType.Name
5: RICE_Supported___c
6: CreatedDate
7: AccountId
8: Lead_Faculty___c
9: Parent_Opportunity___c (factor)
10: RecordType.Name.1
11: Industry
12: Business_Type___c
13: ParentId (factor)
14: RecordType
15: CreatedDate_month
16: Is_External___c

```r
# Reading Data In
data <- read.csv("/Users/ChantelleChiu/Documents/GitHub/Project-90106-G28/Chantelle/clea
                 header = TRUE, stringsAsFactors = T)
names(data) <- c("X1", "X2", "X3", "X4", "X5", "X6", "X7", "X8", "X9", "X10",
                 "X11", "X12", "X13", "X14", "X15", "X16")

#summary(data)
```

Build regression model with 1 predictor ("StageName") and 14 features (X3-X16)

```r
mod1 <- glm(factor(X2)~X3+X4+X5+X6+X8+factor(X9)+X10+X11+X12+factor(X13)+X14+X15+X16,
            family = "binomial", data = data)

# 10-fold cross_validation
train_control <- trainControl(method = "cv", number = 10)
```

```
# train the model on data set
model <- train(factor(X2)~X3+X4+X5+X6+X8+factor(X9)+X10+X11+X12+factor(X13)+X14+X15+X16
               data = data,
               trControl = train_control,
               method = "glm",
               family="binomial")

# print cv scores
print(model)

## Generalized Linear Model
##
## 6052 samples
##   13 predictor
##    2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 5446, 5447, 5447, 5447, 5446, 5447, ...
## Resampling results:
##
##   Accuracy   Kappa
##   0.9983477  0.9963151

# summary(model)
```

Conclusion:

Accuracy 99% No attributes are significant (p-value < 0.05)

Next step:

Remove feature X3: "Status_Reason___c" Build regression model with 1 predictor ("Stage-Name") and 13 features (X4-X16)

```
mod2 <- glm(factor(X2)~X4+X5+X6+X8+factor(X9)+X10+X11+X12+factor(X13)+X14+X15+X16,
            family = "binomial", data = data)

# 10-fold cross_validation
train_control <- trainControl(method = "cv", number = 10)

# train the model on data set
model <- train(factor(X2)~X4+X5+X6+X8+factor(X9)+X10+X11+X12+factor(X13)+X14+X15+X16,
               data = data,
               trControl = train_control,
               method = "glm",
```

```
                  family="binomial")
```

```
summary(model)
```

```
##
## Call:
## NULL
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -2.5766  -0.8521  -0.3851   0.9455   3.1094
##
## Coefficients: (11 not defined because of singularities)
##                                                    Estimate
## (Intercept)                                       -7.353e+12
## `X4Competitive Bid`                               -3.844e+00
## `X4Consultancy (Non-research)`                    -2.027e+00
## `X4Custom Education (MSPACE Included)`             1.449e+13
## X4Grants                                          -1.657e+00
## X4Internship                                       8.799e+13
## X4Panel                                           -1.895e+00
## `X4Parent Grant`                                   4.917e+00
## X4Philanthropic                                   -1.761e+00
## `X4Research Contract`                             -1.624e+00
## X4Sponsorship                                     -2.073e+00
## `X4Technology Transfer`                           -2.103e+00
## `X5RIC RE&D and BD&I`                              2.207e+00
## `X5RIC-BD&I`                                       1.278e+00
## `X5RIC-RE&D`                                       7.412e-01
## X6                                                 4.835e-01
## X80012e000002ZFZYAA4                               3.722e+00
## X80012e000002ZGfbAAG                               3.583e+00
## X80012e000002ZLg7AAG                               3.415e+00
## X80012e000002ZLgIAAW                               2.326e+00
## X80012e000002ZLgJAAW                               3.250e+00
## X80012e000002ZmnxAAC                               2.308e+00
## X80012e000002ZmnzAAC                               2.881e+00
## X80012e000002Zmo0AAC                               4.339e+00
## X80012e000002ZmOUAA0                               3.243e+00
## X80012e000002ZmOZAA0                               3.470e+00
## X80012e000002ZNYOAA4                               3.174e+00
## X80012e000002Zt0MAAS                              -8.082e+00
## X80012e000002Zt1iAAC                              -4.504e+15
## X80012e000002Zt1RAAS                               3.285e+00
## X80012e000002Zt1zAAC                              -5.232e+01
```

```
## `factor(X9)1`                                                               -4.574e-01
## `X10Business Organization`                                                    7.353e+12
## `X10University Department`                                                     7.353e+12
## `X11Agriculture & Horticulture`                                              -1.151e+00
## `X11Aquaculture & Fisheries`                                                 -8.415e-01
## `X11Construction & Construction Materials`                                   -9.322e-01
## X11Defence                                                                   -3.494e-01
## X11Education                                                                          NA
## `X11Energy & Utilities`                                                      -8.986e-01
## `X11Forestry & Paper`                                                         5.942e-01
## X11Health                                                                    -9.888e-01
## `X11Manufacturing & R&D`                                                     -1.542e+00
## `X11Materials: Chemicals & Industrial Metals`                               -1.666e+00
## `X11Mining & Refining`                                                       -1.443e+00
## `X11Mining Equipment and Technology Services`                               -2.077e+00
## `X11Public Administration`                                                   -1.318e+00
## X11Services                                                                  -1.564e+00
## `X11Technology: IT & Telephone Services`                                     -1.367e+00
## X11Transportation                                                            -2.748e+01
## `X11Transportation & Logistics`                                              -1.516e+00
## X11Veterinary                                                                -4.393e-01
## `X12Government: Australia: Federal`                                          -3.240e-01
## `X12Government: Australia: Local & State`                                    -3.195e-01
## `X12Government: International`                                               -6.884e-01
## `X12Large Australian Corporate`                                             -5.346e-01
## `X12Multinational / Other Large Corporate`                                  -1.061e+00
## `X12Not for profit`                                                         -9.517e-01
## `X12PFRO (Publicly-Funded Research Organisation)`                           -1.748e+00
## `X12RDC (Rural Research and Development Corporations)`                       -7.638e-01
## `X12SME (small to medium enterprise)`                                       -6.085e-01
## X12University                                                                8.324e-01
## X13                                                                         -1.703e-01
## `factor(X14)Competitive BidBusiness Organization`                                    NA
## `factor(X14)Consultancy (Non-research)Administrative`                        7.353e+12
## `factor(X14)Consultancy (Non-research)Business Organization`                 3.112e-02
## `factor(X14)Consultancy (Non-research)University Department`                         NA
## `factor(X14)Custom Education (MSPACE Included)Business Organization` -1.449e+13
## `factor(X14)Custom Education (MSPACE Included)University Department` -1.449e+13
## `factor(X14)GrantsBusiness Organization`                                    -1.106e+00
## `factor(X14)GrantsUniversity Department`                                             NA
## `factor(X14)InternshipBusiness Organization`                                -8.799e+13
## `factor(X14)InternshipUniversity Department`                                -8.799e+13
## `factor(X14)PanelBusiness Organization`                                              NA
## `factor(X14)Parent GrantBusiness Organization`                                       NA
## `factor(X14)PhilanthropicBusiness Organization`                                      NA
```

```
## `factor(X14)Research ContractBusiness Organization`           -7.641e-01
## `factor(X14)Research ContractUniversity Department`                    NA
## `factor(X14)SponsorshipBusiness Organization`                          NA
## `factor(X14)Technology TransferBusiness Organization`                  NA
## `factor(X14)Technology TransferUniversity Department`                  NA
## X15                                                             1.720e-02
##                                                                Std. Error
## (Intercept)                                                     9.236e+12
## `X4Competitive Bid`                                             8.157e-01
## `X4Consultancy (Non-research)`                                  1.709e+00
## `X4Custom Education (MSPACE Included)`                           1.076e+14
## X4Grants                                                        1.692e+00
## X4Internship                                                    1.636e+14
## X4Panel                                                         7.442e-01
## `X4Parent Grant`                                                1.245e+00
## X4Philanthropic                                                 7.826e-01
## `X4Research Contract`                                           1.389e+00
## X4Sponsorship                                                   8.040e-01
## `X4Technology Transfer`                                         7.004e-01
## `X5RIC RE&D and BD&I`                                           2.029e-01
## `X5RIC-BD&I`                                                    1.898e-01
## `X5RIC-RE&D`                                                    2.591e-01
## X6                                                              4.112e-02
## X80012e000002ZFZYAA4                                            7.303e-01
## X80012e000002ZGfbAAG                                            2.526e-01
## X80012e000002ZLg7AAG                                            2.530e-01
## X80012e000002ZLgIAAW                                            3.064e-01
## X80012e000002ZLgJAAW                                            2.756e-01
## X80012e000002ZmnxAAC                                            2.606e-01
## X80012e000002ZmnzAAC                                            2.857e-01
## X80012e000002Zmo0AAC                                            2.553e-01
## X80012e000002ZmOUAA0                                            9.372e-01
## X80012e000002ZmOZAA0                                            2.841e-01
## X80012e000002ZNYOAA4                                            2.743e-01
## X80012e000002Zt0MAAS                                            1.442e+02
## X80012e000002Zt1iAAC                                            6.711e+07
## X80012e000002Zt1RAAS                                            3.051e-01
## X80012e000002Zt1zAAC                                            6.711e+07
## `factor(X9)1`                                                   1.496e-01
## `X10Business Organization`                                      9.236e+12
## `X10University Department`                                      9.236e+12
## `X11Agriculture & Horticulture`                                 1.678e+00
## `X11Aquaculture & Fisheries`                                    2.061e+00
## `X11Construction & Construction Materials`                      1.682e+00
## X11Defence                                                      1.673e+00
```

```
## X11Education                                                              NA
## `X11Energy & Utilities`                                            1.677e+00
## `X11Forestry & Paper`                                              2.047e+00
## X11Health                                                          1.668e+00
## `X11Manufacturing & R&D`                                           1.680e+00
## `X11Materials: Chemicals & Industrial Metals`                      1.717e+00
## `X11Mining & Refining`                                             1.699e+00
## `X11Mining Equipment and Technology Services`                      1.841e+00
## `X11Public Administration`                                         1.671e+00
## X11Services                                                        1.668e+00
## `X11Technology: IT & Telephone Services`                           1.678e+00
## X11Transportation                                                  2.980e+05
## `X11Transportation & Logistics`                                    1.692e+00
## X11Veterinary                                                      1.714e+00
## `X12Government: Australia: Federal`                                 4.227e-01
## `X12Government: Australia: Local & State`                          4.202e-01
## `X12Government: International`                                      4.506e-01
## `X12Large Australian Corporate`                                    4.302e-01
## `X12Multinational / Other Large Corporate`                         4.242e-01
## `X12Not for profit`                                                4.214e-01
## `X12PFRO (Publicly-Funded Research Organisation)`                  5.424e-01
## `X12RDC (Rural Research and Development Corporations)`             5.947e-01
## `X12SME (small to medium enterprise)`                              4.209e-01
## X12University                                                      4.380e-01
## X13                                                                1.008e-01
## `factor(X14)Competitive BidBusiness Organization`                         NA
## `factor(X14)Consultancy (Non-research)Administrative`              9.236e+12
## `factor(X14)Consultancy (Non-research)Business Organization`       1.565e+00
## `factor(X14)Consultancy (Non-research)University Department`              NA
## `factor(X14)Custom Education (MSPACE Included)Business Organization` 1.076e+14
## `factor(X14)Custom Education (MSPACE Included)University Department` 1.076e+14
## `factor(X14)GrantsBusiness Organization`                           1.557e+00
## `factor(X14)GrantsUniversity Department`                                  NA
## `factor(X14)InternshipBusiness Organization`                       1.636e+14
## `factor(X14)InternshipUniversity Department`                       1.636e+14
## `factor(X14)PanelBusiness Organization`                                   NA
## `factor(X14)Parent GrantBusiness Organization`                            NA
## `factor(X14)PhilanthropicBusiness Organization`                           NA
## `factor(X14)Research ContractBusiness Organization`                1.208e+00
## `factor(X14)Research ContractUniversity Department`                       NA
## `factor(X14)SponsorshipBusiness Organization`                             NA
## `factor(X14)Technology TransferBusiness Organization`                     NA
## `factor(X14)Technology TransferUniversity Department`                     NA
## X15                                                                1.044e-02
##                                                                      z value
```

```
## (Intercept)                                      -7.960e-01
## `X4Competitive Bid`                              -4.713e+00
## `X4Consultancy (Non-research)`                   -1.186e+00
## `X4Custom Education (MSPACE Included)`            1.350e-01
## X4Grants                                         -9.800e-01
## X4Internship                                     5.380e-01
## X4Panel                                          -2.546e+00
## `X4Parent Grant`                                 3.949e+00
## X4Philanthropic                                  -2.250e+00
## `X4Research Contract`                            -1.170e+00
## X4Sponsorship                                    -2.579e+00
## `X4Technology Transfer`                          -3.003e+00
## `X5RIC RE&D and BD&I`                            1.088e+01
## `X5RIC-BD&I`                                     6.733e+00
## `X5RIC-RE&D`                                     2.861e+00
## X6                                               1.176e+01
## X80012e000002ZFZYAA4                             5.096e+00
## X80012e000002ZGfbAAG                             1.419e+01
## X80012e000002ZLg7AAG                             1.350e+01
## X80012e000002ZLgIAAW                             7.593e+00
## X80012e000002ZLgJAAW                             1.179e+01
## X80012e000002ZmnxAAC                             8.857e+00
## X80012e000002ZmnzAAC                             1.008e+01
## X80012e000002Zmo0AAC                             1.700e+01
## X80012e000002Zm0UAA0                             3.460e+00
## X80012e000002Zm0ZAA0                             1.221e+01
## X80012e000002ZNYOAA4                             1.157e+01
## X80012e000002Zt0MAAS                             -5.600e-02
## X80012e000002Zt1iAAC                             -6.711e+07
## X80012e000002Zt1RAAS                             1.077e+01
## X80012e000002Zt1zAAC                             0.000e+00
## `factor(X9)1`                                    -3.057e+00
## `X10Business Organization`                       7.960e-01
## `X10University Department`                        7.960e-01
## `X11Agriculture & Horticulture`                  -6.860e-01
## `X11Aquaculture & Fisheries`                     -4.080e-01
## `X11Construction & Construction Materials`       -5.540e-01
## X11Defence                                       -2.090e-01
## X11Education                                               NA
## `X11Energy & Utilities`                          -5.360e-01
## `X11Forestry & Paper`                            2.900e-01
## X11Health                                        -5.930e-01
## `X11Manufacturing & R&D`                         -9.180e-01
## `X11Materials: Chemicals & Industrial Metals`    -9.710e-01
## `X11Mining & Refining`                           -8.500e-01
```

```
## `X11Mining Equipment and Technology Services`                                    -1.128e+00
## `X11Public Administration`                                                        -7.890e-01
## X11Services                                                                       -9.380e-01
## `X11Technology: IT & Telephone Services`                                          -8.150e-01
## X11Transportation                                                                  0.000e+00
## `X11Transportation & Logistics`                                                    -8.960e-01
## X11Veterinary                                                                      -2.560e-01
## `X12Government: Australia: Federal`                                                -7.660e-01
## `X12Government: Australia: Local & State`                                          -7.600e-01
## `X12Government: International`                                                     -1.528e+00
## `X12Large Australian Corporate`                                                    -1.243e+00
## `X12Multinational / Other Large Corporate`                                         -2.501e+00
## `X12Not for profit`                                                                -2.259e+00
## `X12PFRO (Publicly-Funded Research Organisation)`                                  -3.222e+00
## `X12RDC (Rural Research and Development Corporations)`                             -1.284e+00
## `X12SME (small to medium enterprise)`                                              -1.446e+00
## X12University                                                                       1.901e+00
## X13                                                                                -1.689e+00
## `factor(X14)Competitive BidBusiness Organization`                                         NA
## `factor(X14)Consultancy (Non-research)Administrative`                               7.960e-01
## `factor(X14)Consultancy (Non-research)Business Organization`                        2.000e-02
## `factor(X14)Consultancy (Non-research)University Department`                              NA
## `factor(X14)Custom Education (MSPACE Included)Business Organization` -1.350e-01
## `factor(X14)Custom Education (MSPACE Included)University Department` -1.350e-01
## `factor(X14)GrantsBusiness Organization`                                           -7.110e-01
## `factor(X14)GrantsUniversity Department`                                                  NA
## `factor(X14)InternshipBusiness Organization`                                       -5.380e-01
## `factor(X14)InternshipUniversity Department`                                       -5.380e-01
## `factor(X14)PanelBusiness Organization`                                                   NA
## `factor(X14)Parent GrantBusiness Organization`                                            NA
## `factor(X14)PhilanthropicBusiness Organization`                                           NA
## `factor(X14)Research ContractBusiness Organization`                                -6.330e-01
## `factor(X14)Research ContractUniversity Department`                                       NA
## `factor(X14)SponsorshipBusiness Organization`                                             NA
## `factor(X14)Technology TransferBusiness Organization`                                     NA
## `factor(X14)Technology TransferUniversity Department`                                     NA
## X15                                                                                 1.648e+00
##                                                                                    Pr(>|z|)
## (Intercept)                                                                        0.425948
## `X4Competitive Bid`                                                                2.44e-06
## `X4Consultancy (Non-research)`                                                     0.235747
## `X4Custom Education (MSPACE Included)`                                             0.892896
## X4Grants                                                                           0.327322
## X4Internship                                                                       0.590629
## X4Panel                                                                            0.010896
```

```
## `X4Parent Grant`                                          7.86e-05
## X4Philanthropic                                           0.024461
## `X4Research Contract`                                      0.242183
## X4Sponsorship                                             0.009917
## `X4Technology Transfer`                                    0.002673
## `X5RIC RE&D and BD&I`                                       < 2e-16
## `X5RIC-BD&I`                                               1.66e-11
## `X5RIC-RE&D`                                               0.004221
## X6                                                          < 2e-16
## X80012e000002ZFZYAA4                                       3.47e-07
## X80012e000002ZGfbAAG                                        < 2e-16
## X80012e000002ZLg7AAG                                        < 2e-16
## X80012e000002ZLgIAAW                                       3.12e-14
## X80012e000002ZLgJAAW                                        < 2e-16
## X80012e000002ZmnxAAC                                        < 2e-16
## X80012e000002ZmnzAAC                                        < 2e-16
## X80012e000002Zmo0AAC                                        < 2e-16
## X80012e000002ZmOUAA0                                       0.000539
## X80012e000002ZmOZAA0                                        < 2e-16
## X80012e000002ZNYOAA4                                        < 2e-16
## X80012e000002Zt0MAAS                                       0.955301
## X80012e000002Zt1iAAC                                        < 2e-16
## X80012e000002Zt1RAAS                                        < 2e-16
## X80012e000002Zt1zAAC                                       0.999999
## `factor(X9)1`                                              0.002233
## `X10Business Organization`                                 0.425948
## `X10University Department`                                 0.425948
## `X11Agriculture & Horticulture`                            0.492660
## `X11Aquaculture & Fisheries`                               0.683093
## `X11Construction & Construction Materials`                 0.579438
## X11Defence                                                 0.834596
## X11Education                                                     NA
## `X11Energy & Utilities`                                    0.592138
## `X11Forestry & Paper`                                      0.771609
## X11Health                                                  0.553288
## `X11Manufacturing & R&D`                                   0.358693
## `X11Materials: Chemicals & Industrial Metals`              0.331783
## `X11Mining & Refining`                                     0.395524
## `X11Mining Equipment and Technology Services`              0.259282
## `X11Public Administration`                                 0.430177
## X11Services                                                0.348233
## `X11Technology: IT & Telephone Services`                   0.415149
## X11Transportation                                          0.999926
## `X11Transportation & Logistics`                            0.370319
## X11Veterinary                                              0.797788
```

```
## `X12Government: Australia: Federal`                                              0.443401
## `X12Government: Australia: Local & State`                                        0.447045
## `X12Government: International`                                                   0.126553
## `X12Large Australian Corporate`                                                 0.214013
## `X12Multinational / Other Large Corporate`                                      0.012395
## `X12Not for profit`                                                             0.023902
## `X12PFRO (Publicly-Funded Research Organisation)`                               0.001272
## `X12RDC (Rural Research and Development Corporations)`                           0.199026
## `X12SME (small to medium enterprise)`                                           0.148276
## X12University                                                                   0.057363
## X13                                                                             0.091290
## `factor(X14)Competitive BidBusiness Organization`                                     NA
## `factor(X14)Consultancy (Non-research)Administrative`                           0.425948
## `factor(X14)Consultancy (Non-research)Business Organization`                    0.984136
## `factor(X14)Consultancy (Non-research)University Department`                           NA
## `factor(X14)Custom Education (MSPACE Included)Business Organization` 0.892896
## `factor(X14)Custom Education (MSPACE Included)University Department` 0.892896
## `factor(X14)GrantsBusiness Organization`                                        0.477315
## `factor(X14)GrantsUniversity Department`                                              NA
## `factor(X14)InternshipBusiness Organization`                                    0.590629
## `factor(X14)InternshipUniversity Department`                                    0.590629
## `factor(X14)PanelBusiness Organization`                                               NA
## `factor(X14)Parent GrantBusiness Organization`                                       NA
## `factor(X14)PhilanthropicBusiness Organization`                                      NA
## `factor(X14)Research ContractBusiness Organization`                             0.526949
## `factor(X14)Research ContractUniversity Department`                                   NA
## `factor(X14)SponsorshipBusiness Organization`                                        NA
## `factor(X14)Technology TransferBusiness Organization`                                NA
## `factor(X14)Technology TransferUniversity Department`                                NA
## X15                                                                             0.099385
## 
## (Intercept)
## `X4Competitive Bid`                                             ***
## `X4Consultancy (Non-research)`
## `X4Custom Education (MSPACE Included)`
## X4Grants
## X4Internship
## X4Panel                                                        *
## `X4Parent Grant`                                               ***
## X4Philanthropic                                                *
## `X4Research Contract`
## X4Sponsorship                                                  **
## `X4Technology Transfer`                                        **
## `X5RIC RE&D and BD&I`                                          ***
## `X5RIC-BD&I`                                                   ***
```

```
## `X5RIC-RE&D`                                              **
## X6                                                        ***
## X80012e000002ZFZYAA4                                      ***
## X80012e000002ZGfbAAG                                      ***
## X80012e000002ZLg7AAG                                      ***
## X80012e000002ZLgIAAW                                      ***
## X80012e000002ZLgJAAW                                      ***
## X80012e000002ZmnxAAC                                      ***
## X80012e000002ZmnzAAC                                      ***
## X80012e000002Zmo0AAC                                      ***
## X80012e000002ZmOUAA0                                      ***
## X80012e000002ZmOZAA0                                      ***
## X80012e000002ZNYOAA4                                      ***
## X80012e000002Zt0MAAS
## X80012e000002Zt1iAAC                                      ***
## X80012e000002Zt1RAAS                                      ***
## X80012e000002Zt1zAAC
## `factor(X9)1`                                             **
## `X10Business Organization`
## `X10University Department`
## `X11Agriculture & Horticulture`
## `X11Aquaculture & Fisheries`
## `X11Construction & Construction Materials`
## X11Defence
## X11Education
## `X11Energy & Utilities`
## `X11Forestry & Paper`
## X11Health
## `X11Manufacturing & R&D`
## `X11Materials: Chemicals & Industrial Metals`
## `X11Mining & Refining`
## `X11Mining Equipment and Technology Services`
## `X11Public Administration`
## X11Services
## `X11Technology: IT & Telephone Services`
## X11Transportation
## `X11Transportation & Logistics`
## X11Veterinary
## `X12Government: Australia: Federal`
## `X12Government: Australia: Local & State`
## `X12Government: International`
## `X12Large Australian Corporate`
## `X12Multinational / Other Large Corporate`                *
## `X12Not for profit`                                       *
## `X12PFRO (Publicly-Funded Research Organisation)`         **
```

```
## `X12RDC (Rural Research and Development Corporations)`
## `X12SME (small to medium enterprise)`
## X12University                                                          .
## X13                                                                    .
## `factor(X14)Competitive BidBusiness Organization`
## `factor(X14)Consultancy (Non-research)Administrative`
## `factor(X14)Consultancy (Non-research)Business Organization`
## `factor(X14)Consultancy (Non-research)University Department`
## `factor(X14)Custom Education (MSPACE Included)Business Organization`
## `factor(X14)Custom Education (MSPACE Included)University Department`
## `factor(X14)GrantsBusiness Organization`
## `factor(X14)GrantsUniversity Department`
## `factor(X14)InternshipBusiness Organization`
## `factor(X14)InternshipUniversity Department`
## `factor(X14)PanelBusiness Organization`
## `factor(X14)Parent GrantBusiness Organization`
## `factor(X14)PhilanthropicBusiness Organization`
## `factor(X14)Research ContractBusiness Organization`
## `factor(X14)Research ContractUniversity Department`
## `factor(X14)SponsorshipBusiness Organization`
## `factor(X14)Technology TransferBusiness Organization`
## `factor(X14)Technology TransferUniversity Department`
## X15                                                                    .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 7750.2  on 6051  degrees of freedom
## Residual deviance: 6079.3  on 5981  degrees of freedom
## AIC: 6221.3
##
## Number of Fisher Scoring iterations: 25
```

Conclusion:

Recall the attributes:

1: Id
2: StageName (factor)
3: Status_Reason___c
4: RecordType.Name
5: RICE_Supported___c
6: CreatedDate
7: AccountId

8: Lead_Faculty___c
9: Parent_Opportunity___c (factor)
10: RecordType.Name.1
11: Industry
12: Business_Type___c
13: ParentId (factor)
14: RecordType
15: CreatedDate_month
16: Is_External___c

Significant Attributes:

Very significant (p-value < 0.001):

X4Competitive Bid
X4Parent Grant
X5RIC RE&D and BD&I
X5RIC-BD&I
X6
X80012e000002ZFZYAA4
X80012e000002ZGfbAAG
X80012e000002ZLg7AAG
X80012e000002ZLgIAAW
X80012e000002ZLgJAAW
X80012e000002ZmnxAAC
X80012e000002ZmnzAAC
X80012e000002Zmo0AAC
X80012e000002ZmOUAA0
X80012e000002ZmOZAA0
X80012e000002ZNYOAA4 X80012e000002Zt1iAAC X80012e000002Zt1RAAS

Significant (0,001 < p-value < 0.01)

X4Sponsorship
X4Technology Transfer
X5RIC-RE&D
factor(X9)1
X12PFRO (Publicly-Funded Research Organisation)

Somehow significant (0.01 < p-value < 0.05)

X4Panel
X4Philanthropic
X4Sponsorship
X12Multinational / Other Large Corporate
X12Not for profit

0.05 < p-value < 0.1 (keep or not?)

X12University
X13
X15

```
print(model)
```

```
## Generalized Linear Model
##
## 6052 samples
##   11 predictor
##    2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 5447, 5447, 5447, 5447, 5446, 5447, ...
## Resampling results:
##
##   Accuracy   Kappa
##   0.7219205  0.3716663
```

Conclusion

1. Accuracy drops from 99% to around 73%.
  2. Since we have about 2:1 split for classes 0 and 1 of the "stageName", which is sort of imbalance, we might also want to look at the Kappa accuracy which is nearly 38%. According to the guideline for Kappa score, 38% is a high "fair".

Note

from website https://machinelearningmastery.com/machine-learning-evaluation-metrics-in-r/

Accuracy

is the percentage of correctly classifies instances out of all instances. It is more useful on a binary classification than multi-class classification problems because it can be less clear exactly how the accuracy breaks down across those classes.

Kappa or Cohen's Kappa

is like classification accuracy, except that it is normalized at the baseline of random chance on your dataset. It is a more useful measure to use on problems that have an imbalance in the classes (e.g. 70-30 split for classes 0 and 1 and you can achieve 70% accuracy by predicting all instances are for class 0).

Suggested guidelines for Kappa:

values $<= 0$ as indicating no agreement
0.01–0.20 as none to slight
0.21–0.40 as fair, 0.41– 0.60 as moderate,
0.61–0.80 as substantial
0.81–1.00 as almost perfect agreement.