PGDDS:HR ANAYTICS CASE STUDY

Date:25/11/2018

Team Members

1.Gaurav Sacheva
2.Aman Raj

Business Case

Problem Statement

Every year ~15% of employees leave XYZ company and need to be replaced with the talent pool available in the job market. This percentage of attrition(employees leaving, either on their own or because they got fired) impacts the companynegatively, because of the following reasons:

- The former employees' projects get delayed, which makes it difficult to meet timelines, resulting in a reputation loss among consumers and partners
- A sizeable department has to be maintained, for the purposes of recruiting new talent
- More often than not, the new employees have to be trained for the job and/or given time to acclimatize themselves to the company

Goal

Model the probability of attrition using logistic regression : ¬ To understand what factors XYZ company should focus on to curb attrition, in other words, what changes should be made inthe workplace, inorder

Data Available

5 Data files have been provided for building the model.

- Employee survey data - Provides information on employees mental state
- Manager survey data – Provides information on the performance of an employee under a manager
- In and out time – Gives the time recorded for employee reports in to the office and time at which the employee goes out of the office
- General data – Master data which contains complete details of the employee

Data Cleaning

- NA values in data except in_time and out_time dataframes were identified and were imputed by the median of that column.
- NA columns were identified in in_time and out_time dataframes and deleted
- Columns with same values throughout were deleted.
- Duplication of EmployeeID in various dataframes was also checked.
- Oultlier treatment was also done on Monthly Income Column.Values 1.5 times the interquartile range from third quantile were removed.
- In_time and Out_time data was converted from wide format to long format in-order to calculate the average working hours

Data Preparation

- Binary Categorical were converted in 1 or 0.
- Multivariate Categorical variable were converted to dummy variables and were added to the dataset.
- Continuous variables were scaled.
- Attrition, the target variable is categorical in nature and hence the logistic regression model was used
- Eventually all the data available was merged into a large integrated dataset.
- Attrition rate found in our dataset is 16%(approx) only.
- The data is split into training and testing data in the ratio 7:3

Model Building

- Generalized Linear Model(glm) to build the model to predict attrition.
- Step AIC to get the starting model. Here, the model predicts how Attrition behaves with respect to other variables. Since step AIC suggests the optimal model based on Akaike Information Criteria (model having lowest AIC value is preferred), there are some insignificant variables included in the model which were removed
- VIF (Variable Inflation Factor)>2 criteria was used to check for multicollinearity between the variables. Variables were removed on the basis of this criteria.
- p-value>0.005 criteria was also used to remove insignificant variables from the model.
- In the end 11 variables remained in the model and this model was found after 14 iterations.

MODEL

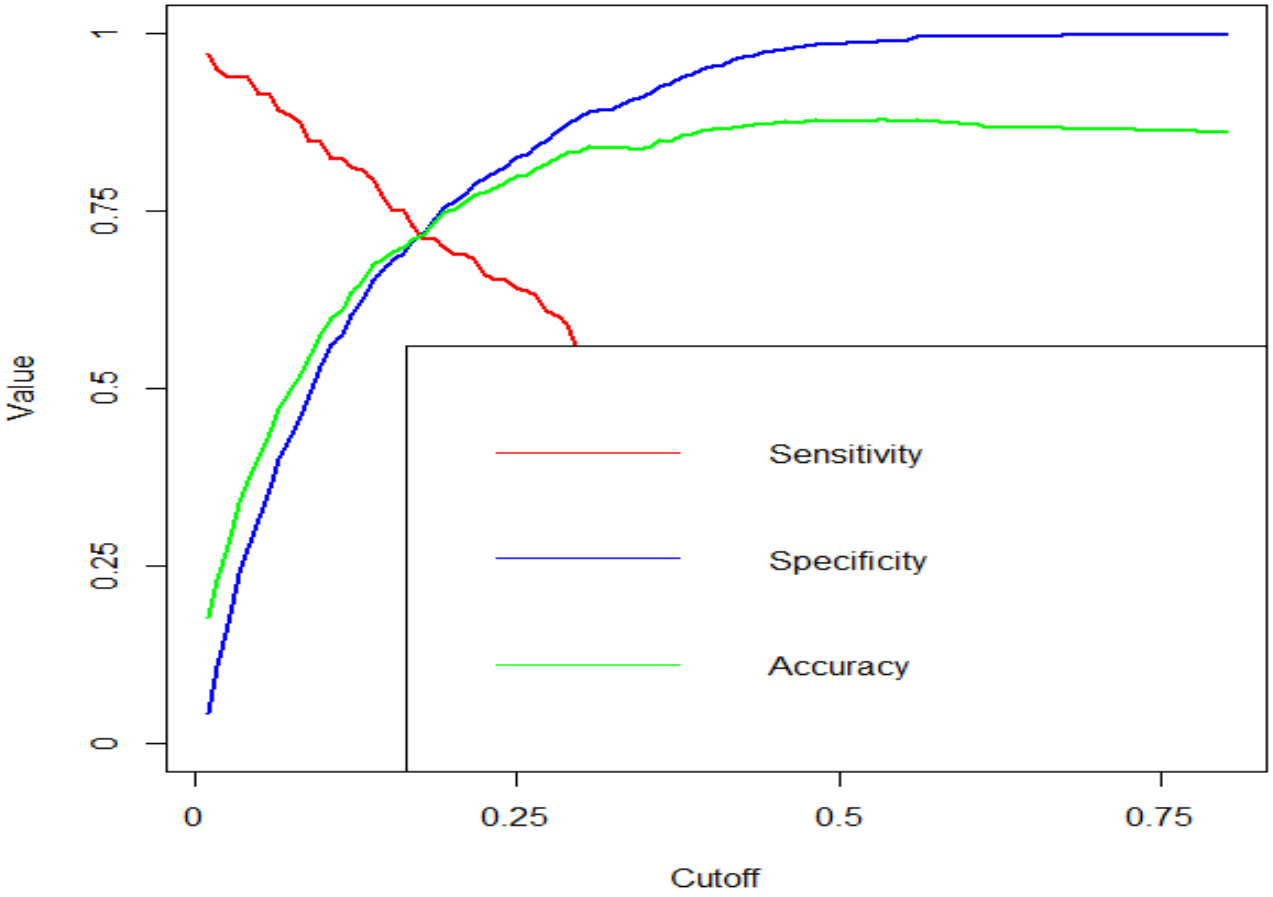
```
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                   0.40574    0.28866   1.406     0.16
Age                          -0.32092    0.07968  -4.028 5.63e-05 ***
NumCompaniesWorked            0.30329    0.05866   5.170 2.34e-07 ***
TotalWorkingYears            -0.57453    0.11059  -5.195 2.05e-07 ***
YearsSinceLastPromotion       0.55559    0.07639   7.273 3.51e-13 ***
YearsWithCurrManager         -0.47372    0.08721  -5.432 5.58e-08 ***
EnvironmentSatisfaction      -0.32879    0.05128  -6.412 1.44e-10 ***
JobSatisfaction              -0.36811    0.05095  -7.225 5.00e-13 ***
WorkLifeBalance              -0.42751    0.07928  -5.392 6.96e-08 ***
AvgWrkHrs                     0.62565    0.05449  11.483  < 2e-16 ***
BusinessTravelTravel_Frequently  0.85176    0.13029   6.537 6.27e-11 ***
MaritalStatusSingle           1.12139    0.11579   9.684  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
```

```
> vif(mdl_14)
                             Age                 NumCompaniesWorked
                        1.774148                           1.238446
                TotalWorkingYears          YearsSinceLastPromotion
                        2.538821                           1.801066
          YearsWithCurrManager           EnvironmentSatisfaction
                        1.771189                           1.039815
                 JobSatisfaction                    WorkLifeBalance
                        1.037157                           1.025596
                     AvgWrkHrs  BusinessTravelTravel_Frequently
                        1.053477                           1.022552
          MaritalStatusSingle
                        1.052804
```

MODEL EVALUATION

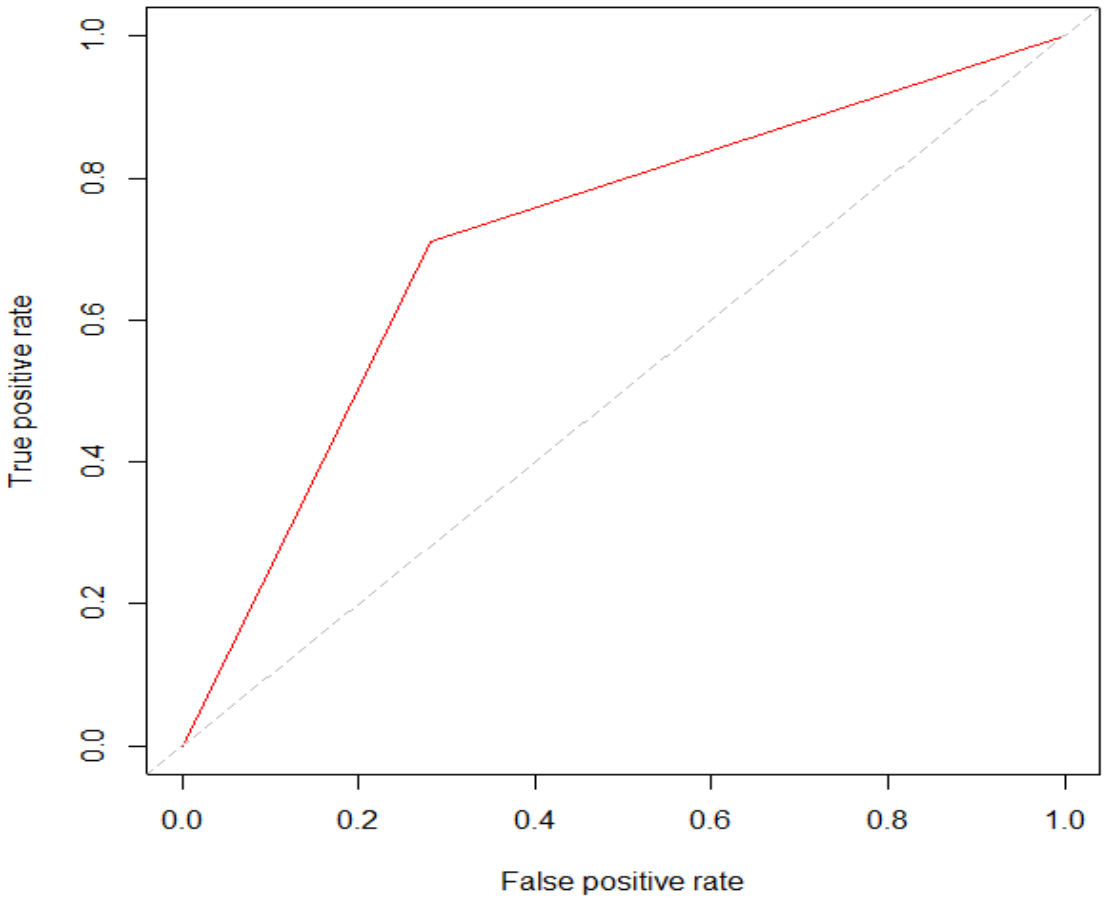
Accuracy,Senstivity,Specificity



CutOff Value:0.178

At Cut off Value
Accuracy : 71.7%
Sensitivity : 71%
Specificity : 71.9%

Area Under Curve(AUC)

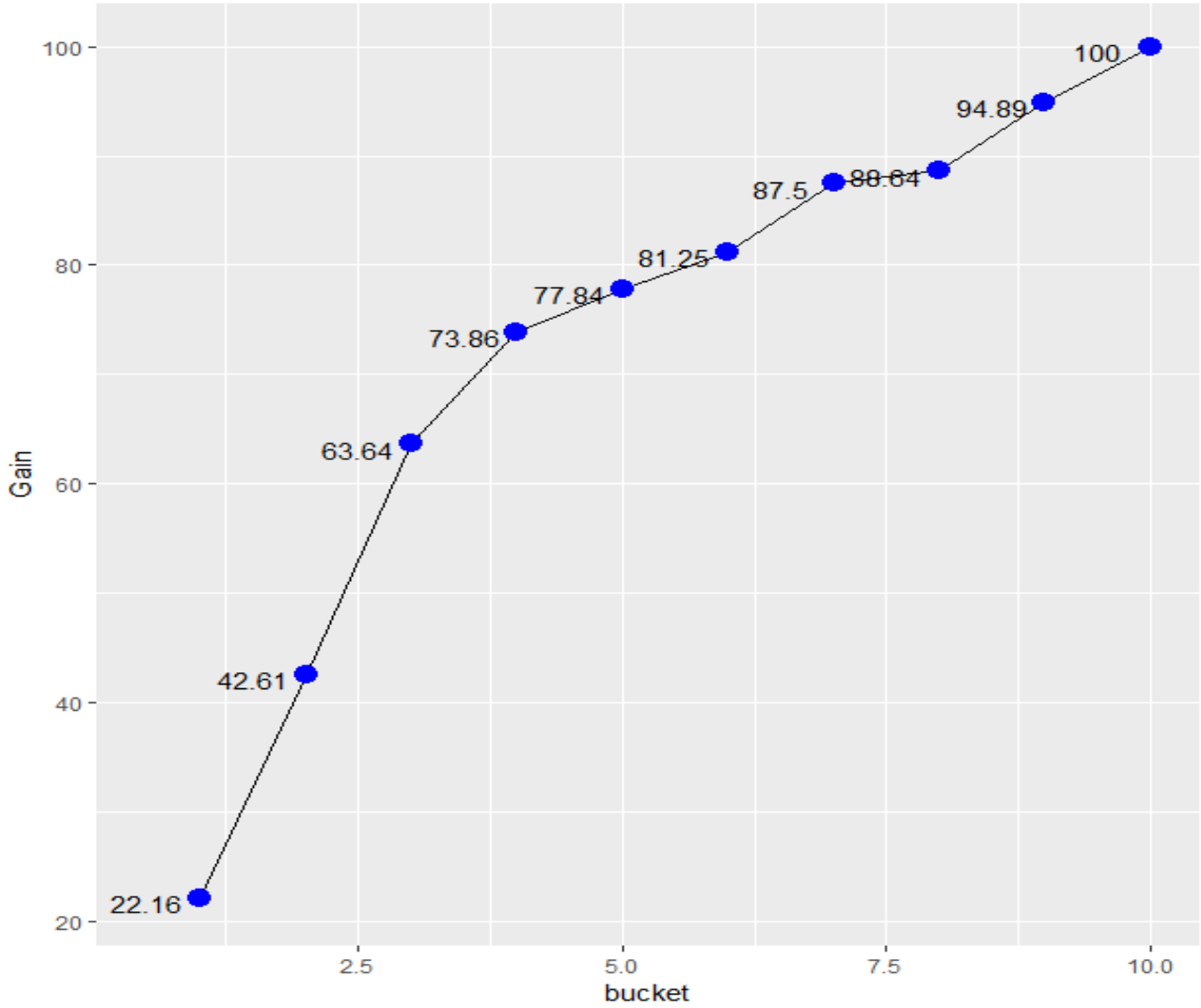


AUC=0.714

Area under the curve is 71.4 as compared to the random model which gives 50%.

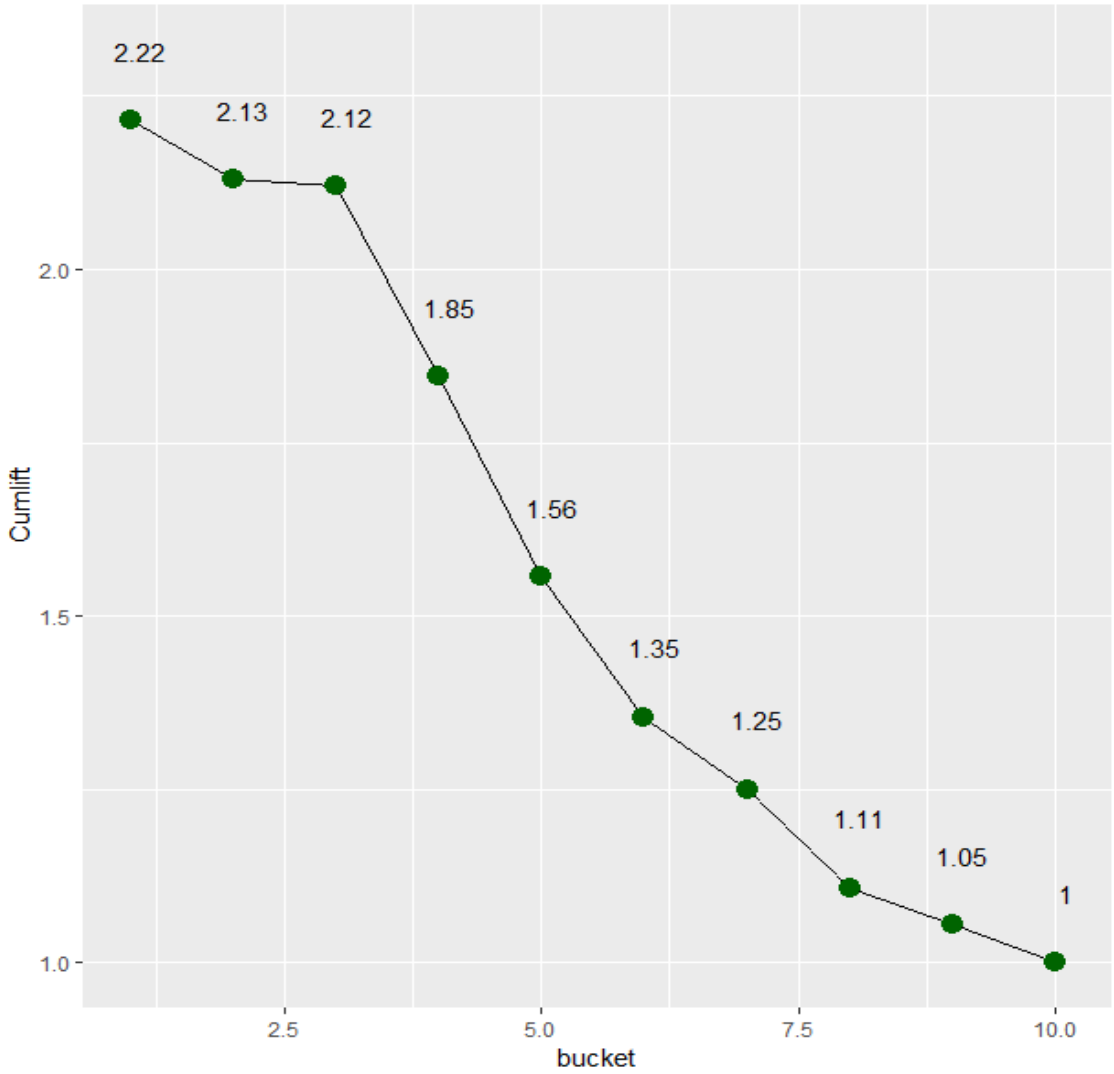
KS Static - 42.9% (KS Static > 40 % indicates a good model).

Gain Chart



Gain is 73.86% at the 4th decile which means if employees are arranged by probability then 73.86 % of the first 40% will be predicted correctly by the model.

Lift Chart



Lift is 2.12 at the3rd decile.This means that the model's gain by the end of the 3rd decile is 2.12 times that of a random model's gain at the end of 3 deciles. In other words, the model catches 2.12 times more attrition than a random model would have caught.

Conclusion

The analysis and model results show that the company should focus on employees with the following attributes to curb attrition :

- Frequent Business Travel
- Single (Marital Status)
- Low Environment Satisfaction
- Low Job Satisfaction
- Low Work Life balance
- Higher average working hours
- Lesser Age(Young people)
- Less Total Working years (Less experience)
- Higher Years since last Promotion
- Less Years with current Manager
- More Number of companies worked(Frequently changing jobs)