PG Diploma in Data Science
March 2018

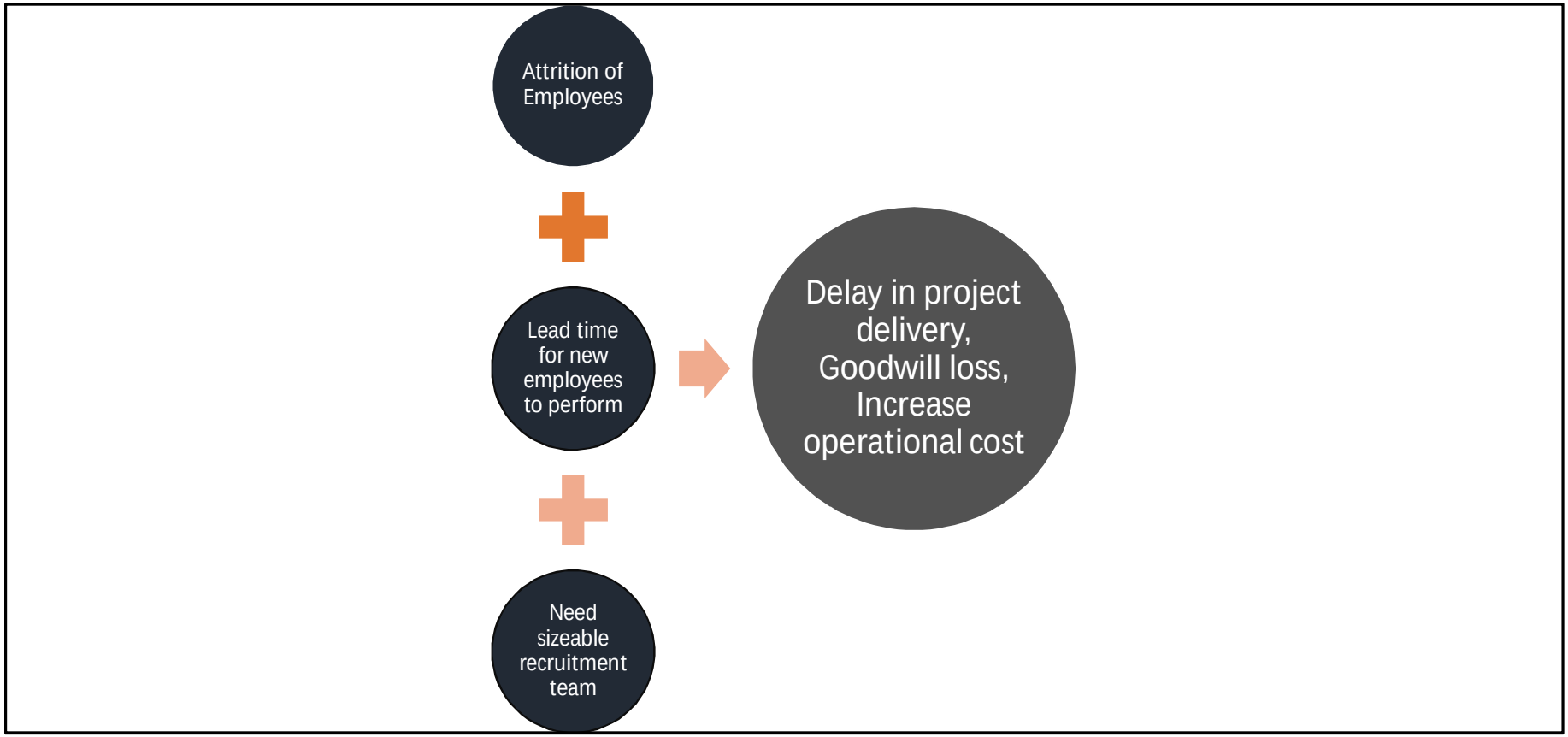
HR Analytics Case Study

AGENDA

- BUSINESS PROBLEM
- BUSINESS OBJECTIVES
- ANALYTICS MODEL DEVELOPMENT STEPS
- DEVELOPING MODEL
- EXPECTED OUTCOME AND INFERENCE
- FUTURE FINE TUNING – BASED ON FEEDBACK

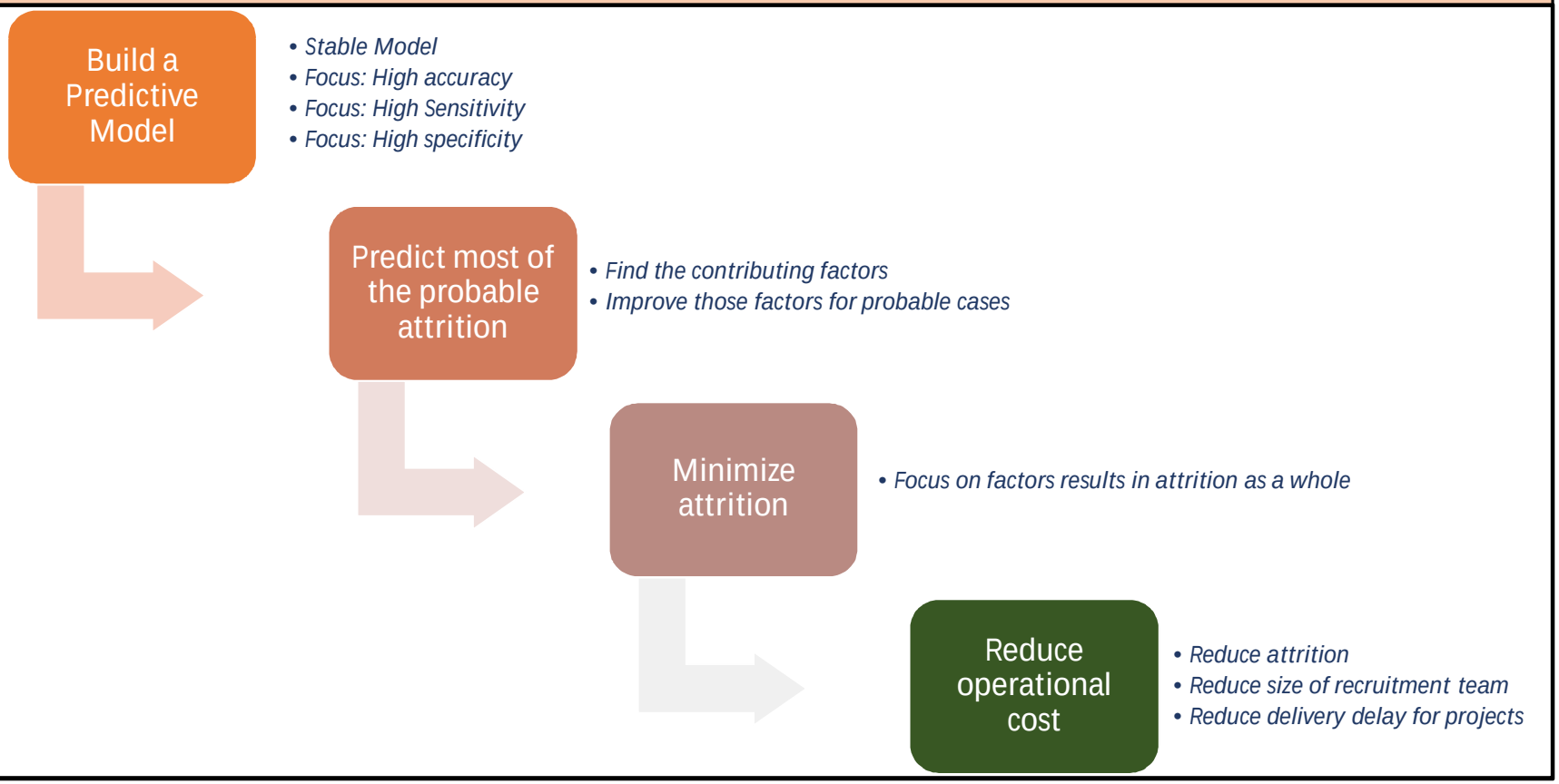
Presented By:
Subhra Sinha
Sathish Karthik Ramu
Saathvik K V
Amitava Mukherjee

HR Analytics Case Study – BUSINESS PROBLEM



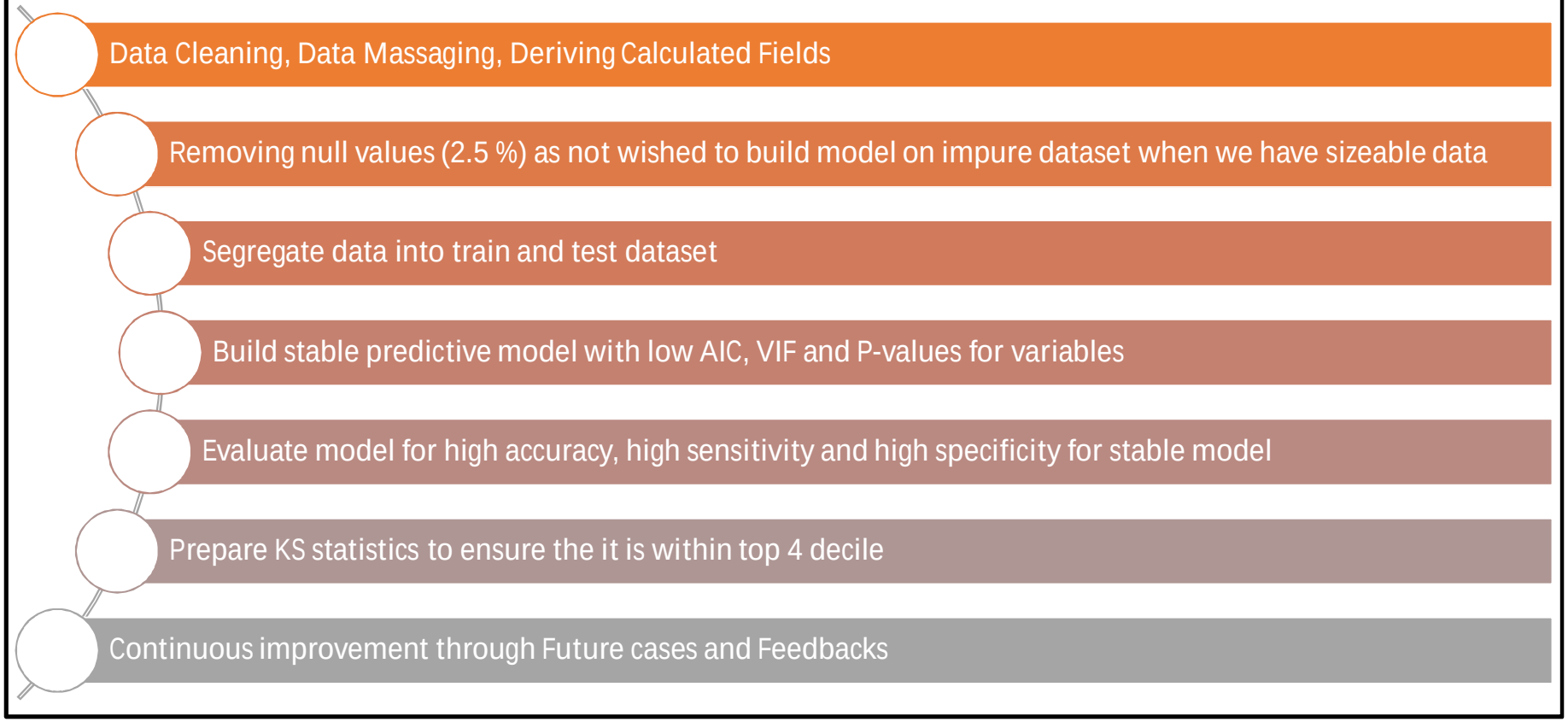
Predictive Analytics Group Case Study – "HR Analytics Case Study"

HR Analytics Case Study – BUSINESS OBJECTIVES

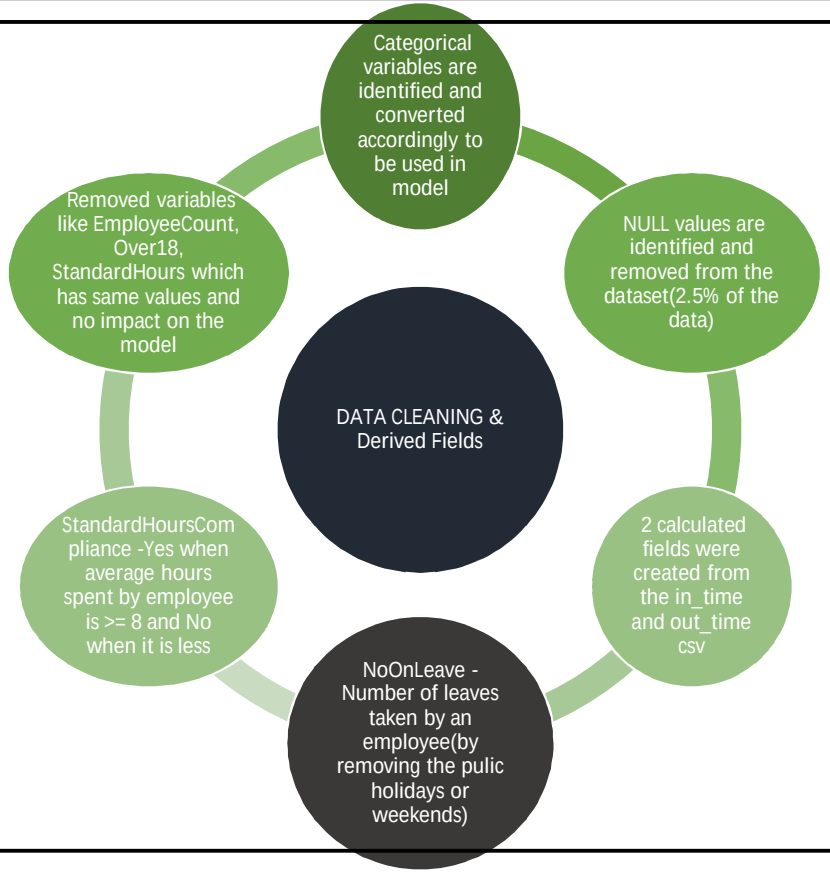


Predictive Analytics Group Case Study – "HR Analytics Case Study"

HR Analytics Case Study – Analytics Model Development Steps



Developing Model – Data cleaning, Derived fields



Developing Model – Low AIC, Low P values, Low VIFs

Build Initial Model consists of all variables

StepAIC in both direction to reduce the number of variables automatically and removed variables whose impact is less or not at all

Build model with low AIC values and variables with low P values and Check for low VIFs

Developing Model – Final Model

```
> summary(model_10)

Call:
glm(formula = Attrition ~ Age + EducationField + NumCompaniesWorked +
    BusinessTravel + TotalWorkingYears + YearsSinceLastPromotion +
    YearsWithCurrManager + EnvironmentSatisfaction + JobSatisfaction +
    WorkLifeBalance + StandardHourCompliance, family = "binomial",
    data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.6750   0.1956   0.3749   0.5809   1.8667

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                     -2.763132   0.538070  -5.135 2.82e-07 ***
Age                              0.040987   0.008667   4.729 2.25e-06 ***
EducationFieldLife Sciences      1.341948   0.326387   4.112 3.93e-05 ***
EducationFieldMarketing          1.494275   0.356999   4.186 2.84e-05 ***
EducationFieldMedical            1.396876   0.329421   4.240 2.23e-05 ***
EducationFieldOther              1.739154   0.402355   4.322 1.54e-05 ***
EducationFieldTechnical Degree   1.715020   0.374148   4.584 4.57e-06 ***
NumCompaniesWorked              -0.143211   0.023680  -6.048 1.47e-09 ***
BusinessTravelTravel_Frequently -1.301225   0.239995  -5.422 5.90e-08 ***
BusinessTravelTravel_Rarely     -0.610724   0.220521  -2.769  0.00561 **
TotalWorkingYears                0.074476   0.013802   5.396 6.81e-08 ***
YearsSinceLastPromotion         -0.177806   0.023398  -7.599 2.98e-14 ***
YearsWithCurrManager             0.129764   0.023948   5.419 6.01e-08 ***
EnvironmentSatisfaction          0.313535   0.051208   6.123 9.19e-10 ***
JobSatisfaction                  0.294001   0.050356   5.838 5.27e-09 ***
WorkLifeBalance                  0.324107   0.077466   4.184 2.87e-05 ***
StandardHourCompliance          -1.557852   0.114002 -13.665  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2661.4  on 3009  degrees of freedom
Residual deviance: 2143.7  on 2993  degrees of freedom
AIC: 2177.7

Number of Fisher Scoring iterations: 5
```

Developing Model – Final Model

Summary Test Predict Data

```
test_pred = predict(final_model, type = "response",
                    newdata = test)


# Let's see the summary

summary(test_pred)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.1229  0.7609  0.8984  0.8359  0.9576  0.9993
```

Confusion Matrix details

```
> test_conf
Confusion Matrix and Statistics

          Reference
Prediction   No  Yes
       No    46   26
       Yes  163 1055

               Accuracy : 0.8535
                 95% CI : (0.833, 0.8724)
    No Information Rate : 0.838
    P-Value [Acc > NIR] : 0.0689

                  Kappa : 0.2665
 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.9759
            Specificity : 0.2201
         Pos Pred Value : 0.8662
         Neg Pred Value : 0.6389
             Prevalence : 0.8380
         Detection Rate : 0.8178
   Detection Prevalence : 0.9442
      Balanced Accuracy : 0.5980

       'Positive' Class : Yes
```

Random probability data

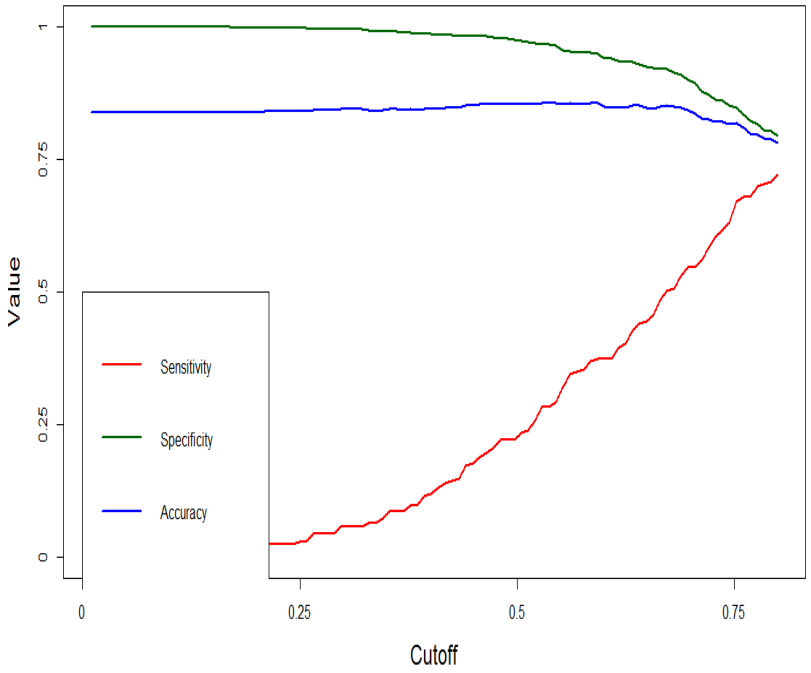
	MaritalStatusMarried	MaritalStatusSingle	prob
0	1	0	0.9277280
0	1	0	0.9514730
0	0	0	0.5500538
0	1	0	0.1751286
0	0	0	0.9320087
1	0	1	0.8959706
0	0	0	0.7316720
0	0	1	0.7953649
0	0	0	0.9610869
0	0	1	0.8417235

Sorted probability data

	MaritalStatusMarried	MaritalStatusSingle	prob
1	1	0	0.9993187
1	1	0	0.9993187
0	1	0	0.9992799
0	1	0	0.9989295
0	0	1	0.9989095
0	0	1	0.9989095
0	1	0	0.9988323
0	1	0	0.9988323
0	0	1	0.9971272
0	0	1	0.9970965
0	0	0	0.9970646

Developing Model – Final Model

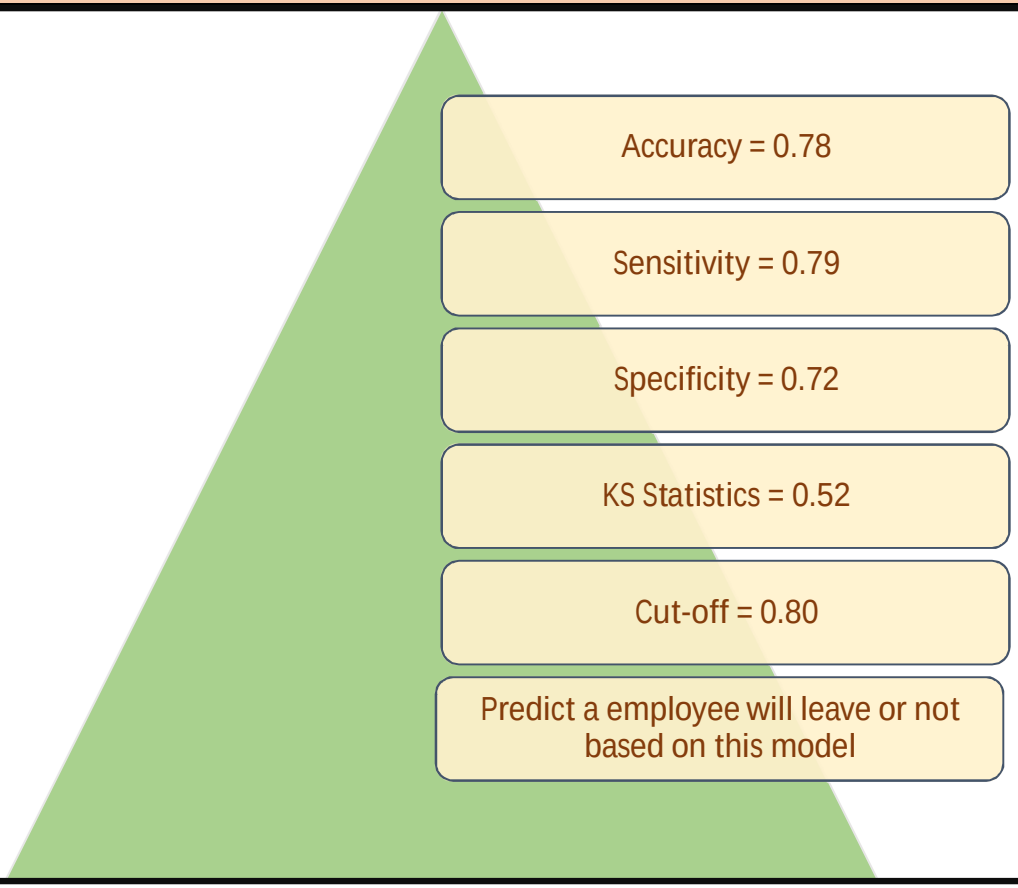
Plot Sensitivity, Specificity and accuracy



Chosen cut-off 0.8 , KS Statistics -> 0.5161976

```
> acc
 Accuracy
0.7821705
>
> sens
Sensitivity
  0.7937095
>
> spec
Specificity
   0.722488
```

EXPECTED OUTCOME AND INFERENCES



FUTURE FINE TUNING – BASED ON FEEDBACK

TEST THE MODEL WITH NEW INPUT

New employees' data

Previously known data

FINE TUNE BASED ON OUTCOME

Match Model outcome Vs. Real outcome

Fine Tune Model to match outcome as close as real outcome

GOAL

- Less Attrition of employees
- Focus on Business travel, Years since last promotion, and the Standard hours compliance to check whether the people will leave

Continuous Fine Tuning

- Fine tune the model in case model outcome is not matching with real outcome
- Fine tune the model in case any new parameter is explored and seems to have major impact