**Highlights of HR\_project.R outputs:**

**Data Structures:**

**data.frame': 14999 obs. of 10 variables:**

**$ satisfaction\_level : num 0.38 0.8 0.11 0.72 0.37 0.41 0.1 0.92 0.89 0.42 ...**

**$ last\_evaluation : num 0.53 0.86 0.88 0.87 0.52 0.5 0.77 0.85 1 0.53 ...**

**$ number\_project : int 2 5 7 5 2 2 6 5 5 2 ...**

**$ average\_montly\_hours : int 157 262 272 223 159 153 247 259 224 142 ...**

**$ time\_spend\_company : int 3 6 4 5 3 3 4 5 5 3 ...**

**$ Work\_accident : int 0 0 0 0 0 0 0 0 0 0 ...**

**$ left : int 1 1 1 1 1 1 1 1 1 1 ...**

**$ promotion\_last\_5years: int 0 0 0 0 0 0 0 0 0 0 ...**

**$ department : chr "sales" "sales" "sales" "sales" ...**

**$ salary : chr "low" "medium" "medium" "low" ...**

Summary of Counts:

accounting hr IT management marketing product\_mng RandD sales support

767 739 1227 630 858 902 787 4140 2229

technical

2720

Chart, bar chart, histogram

Description automatically generated

Chart, bar chart, histogram

Description automatically generated

Proportionately, HR employee has highest percentage of people leaving the job.

A picture containing background pattern

Description automatically generated

Chart, histogram

Description automatically generated

* More employee from Sales, accounting and support left the job.

Chart, bar chart

Description automatically generated

More employees with low salary left the job.

Graphical user interface

Description automatically generated

more than 99% who left did not have promotion last 5 years.

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

**Results from logistic Regression Model:**

Deviance Residuals:

Min 1Q Median 3Q Max

-2.1921 -0.6599 -0.4015 -0.1197 3.0922

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -1.6517677 0.2329999 -7.089 1.35e-12 \*\*\*

satisfaction\_level -4.1197436 0.1178068 -34.970 < 2e-16 \*\*\*

last\_evaluation 0.9042849 0.1782874 5.072 3.94e-07 \*\*\*

number\_project -0.3074209 0.0255463 -12.034 < 2e-16 \*\*\*

average\_montly\_hours 0.0044452 0.0006166 7.210 5.61e-13 \*\*\*

time\_spend\_company 0.2633074 0.0185944 14.161 < 2e-16 \*\*\*

Work\_accident -1.5193812 0.1051285 -14.453 < 2e-16 \*\*\*

promotion\_last\_5years1 -1.5411094 0.3089065 -4.989 6.07e-07 \*\*\*

departmenthr 0.1644849 0.1620927 1.015 0.31022

departmentIT -0.2068839 0.1482290 -1.396 0.16280

departmentmanagement -0.4988573 0.1945119 -2.565 0.01033 \*

departmentmarketing -0.0173623 0.1593772 -0.109 0.91325

departmentproduct\_mng -0.0530144 0.1551137 -0.342 0.73252

departmentRandD -0.4698361 0.1727603 -2.720 0.00654 \*\*

departmentsales 0.0029721 0.1239722 0.024 0.98087

departmentsupport 0.0638583 0.1326189 0.482 0.63015

departmenttechnical 0.1163806 0.1290261 0.902 0.36706

salarylow 1.9675447 0.1526312 12.891 < 2e-16 \*\*\*

salarymedium 1.3830973 0.1536912 8.999 < 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: 11481.4 on 10498 degrees of freedom

Residual deviance: 8963.2 on 10480 degrees of freedom

AIC: 9001.2

Number of Fisher Scoring iterations: 5

Predictedvalues

Actualvalues FALSE TRUE

0 3016 394

1. 537 553

**Model Accuracy** = True Positive + True Negative/Total

= 3016+553/(3016+394+537+553) = 79.31%

**Results from Decision trees Model:**

Diagram, schematic

Description automatically generatedConfusion Matrix and Statistics

predvalues1 No Yes

No 3376 91

Yes 34 999

Accuracy : 0.9722 95% CI : (0.967, 0.9768)

No Information Rate : 0.7578

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.923

Mcnemar's Test P-Value : 5.477e-07

Sensitivity : 0.9900

Specificity : 0.9165

Pos Pred Value : 0.9738

Neg Pred Value : 0.9671

Prevalence : 0.7578

Detection Rate : 0.7502

Detection Prevalence : 0.7704

Balanced Accuracy : 0.9533

'Positive' Class : No

**Results from Random Forests Model:**

RandomForest(formula = trainset$left ~ ., data = trainset, ntree = 100)

Type of random forest: classification

Number of trees: 100

No. of variables tried at each split: 3

OOB estimate of error rate: 0.93%

Confusion matrix:

No Yes class.error

No 7946 17 0.002134874

Yes 81 2445 0.032066508

**Confusion Matrix and Statistics**

predvalues No Yes

No 3457 36

Yes 8 1009

Accuracy : 0.9902

95% CI : (0.9869, 0.9929)

No Information Rate : 0.7683

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.9723

Mcnemar's Test P-Value : 4.693e-05

Sensitivity : 0.9977

Specificity : 0.9656

Pos Pred Value : 0.9897

Neg Pred Value : 0.9921

Prevalence : 0.7683

Detection Rate : 0.7665

Detection Prevalence : 0.7745

Balanced Accuracy : 0.9816

'Positive' Class : No

**List of Important Variables for prediction using Random forests Model:**

**Clearly, you can see that job satisfaction level has greatest impact on making decision whether to continue or leave the job for employees.**

**Please see the plot in next page.**

Table

Description automatically generated

**Results from Naïve Bayes Model:**

**Confusion Matrix and Statistics**

**predvalues No Yes**

**No 3115 346**

**Yes 350 699**

**Accuracy : 0.8457**

**95% CI : (0.8348, 0.8561)**

**No Information Rate : 0.7683**

**P-Value [Acc > NIR] : <2e-16**

**Kappa : 0.5671**

**Mcnemar's Test P-Value : 0.9095**

**Sensitivity : 0.8990**

**Specificity : 0.6689**

**Pos Pred Value : 0.9000**

**Neg Pred Value : 0.6663**

**Prevalence : 0.7683**

**Detection Rate : 0.6907**

**Detection Prevalence : 0.7674**

**Balanced Accuracy : 0.7839**

**'Positive' Class : No**

**Results from Support Vector Machine Model:**

**Call:**

**svm(formula = trainset$left ~ ., data = trainset, kernel = "polynomial", cost = 0.1)**

**Parameters:**

**SVM-Type: C-classification**

**SVM-Kernel: polynomial**

**cost: 0.1**

**degree: 3**

**coef.0: 0**

**Number of Support Vectors: 3637**

**( 1813 1824 )**

**Number of Classes: 2**

**Levels:**

**No Yes**

**Confusion Matrix and Statistics**

**predvalues No Yes**

**No 3369 225**

**Yes 96 820**

**Accuracy : 0.9288**

**95% CI : (0.9209, 0.9362)**

**No Information Rate : 0.7683**

**P-Value [Acc > NIR] : < 2.2e-16**

**Kappa : 0.7911**

**Mcnemar's Test P-Value : 9.048e-13**

**Sensitivity : 0.9723**

**Specificity : 0.7847**

**Pos Pred Value : 0.9374**

**Neg Pred Value : 0.8952**

**Prevalence : 0.7683**

**Detection Rate : 0.7470**

**Detection Prevalence : 0.7969**

**Balanced Accuracy : 0.8785**

**'Positive' Class : No**

**You have seen the performance of different results. One can decide to select model for deployment based on the domain knowledge, accuracy, and consulting with Machine Learning engineers and Dev Ops to see which model will do better job at production level.**

**Thank you.**

**Shobhakhar Adhikari**