Tools used: power Bi, R Studio

Programming Language: R

Data: Employee Attrition Reasons and Survey of Employees

Model used: Logistic Regression

**About Tools:**

Power Bi is an cloud Analytical tool used for data analytics , here we are analyzing data before Prediction and after prediction which are called Power Dashboards .

Step1:

Loading data into RStudio

#Source Path

setwd("C:/Users/vineetha.p.TECHNOVERT/Desktop/Hackathon")

HR\_DataSet<-read.csv("Hr\_Data.csv",header = TRUE,na.strings=c("")) 🡪 data set old data (it consists of employees who left and working , it is used for training model )

Hr\_Prediction\_DataSet<-read.csv("HR\_No.csv",header = TRUE,na.strings=c("")) -> Dataset for doing prediction (currently working employees )

Step2:

# Packages Used

library(caret) -- > package used for partition of data in to Training and testing data

library(sqldf) -> for using SQL Queries in R

step3:

#DataPartition

set.seed(3456)

intrain<-createDataPartition(HR\_DataSet$Attrition,p=0.7,list=FALSE) -> divided data into 2 parts (70% , 30%)

TrainingDataSet<-HR\_DataSet[intrain,] -> 70% data

TestingDataSet<-HR\_DataSet[-intrain,] -> 30% data

Separated data in to parts to train the model (Logistic regression)

Step4:

#Logistic regressionModel

Logistic\_Regression\_Model<-glm(Attrition ~ BusinessTravel+EnvironmentSatisfaction+Gender+MaritalStatus+OverTime+StockOptionLevel+ JobSatisfaction+employeeid+DistanceFromHome+JobLevel+JobRole+RelationshipSatisfaction+Benefits+ExtracurricularActivities,family='binomial',data=TrainingDataSet)

🡪 Training model with Training data using existing columns against attrition

Step5:

#Prediction

PredictedProbability<-predict(Logistic\_Regression\_Model,TestingDataSet,type='response') 🡪 doing prediction with testing data for testing the model

Predicted\_Attrition<-ifelse(PredictedProbability>=0.5,1,0)

Predicted\_Attrition

summary(Logistic\_Regression\_Model) 🡪 forchecking correlation between columns (checking significance of column )

step6:

#Confusion Matrix

ConfusionMatrix<-table(predicted=Predicted\_Attrition, Actual=TestingDataSet$Attrition)

ConfusionMatrix 🡪 helps for calculating accuracy of model

Step7:

#Accuracy

Accuracy<-(sum(diag(ConfusionMatrix))/sum(ConfusionMatrix))\*100

Accuracy 🡪 For Checking accuracy of model

Here accuracy of our model is 85.2%

Step8:

a<-predict(Logistic\_Regression\_Model,Hr\_Prediction\_DataSet,type='response') 🡪 now we are predicting the actual data (Current employees data )

b<- data.frame(empid=Hr\_Prediction\_DataSet$employeeid,a)🡪 probability to leave an employee with employee id

step9:

HR\_Data\_Prediction<-sqldf('

SELECT

employeeid

,Age

,a AS Probability\_to\_Left

,Attrition

,CASE

WHEN a >= 0.5 🡪with probability (>0.5 treat as employee will leave (1))

THEN 1

ELSE 0

END AS Predicted\_Status

,BusinessTravel

,Department

,DistanceFromHome

,EducationField

,EnvironmentSatisfaction

,Gender

,JobInvolvement

,JobLevel

,JobRole

,JobSatisfaction

,MaritalStatus

,NumCompaniesWorked

,Over18

,OverTime

,PercentSalaryHike

,PerformanceRating

,RelationshipSatisfaction

,StockOptionLevel

,TotalWorkingYears

,TrainingTimes

,WorkLifeBalance

,YearsAtCompany

,YearsInCurrentRole

,YearsSinceLastPromotion

,YearsWithCurrManager

,YearOfJoining

,TechnologyAdoption

,Benefits

,Facilities

,ChallengingLevels

,ExtracurricularActivities

,YearOfLeaving

FROM Hr\_Prediction\_DataSet

INNER JOIN b ON Hr\_Prediction\_DataSet.employeeid = b.empid')

HR\_Data\_Prediction 🡪 this data is directly connected to power bi for doing analysis