**Build Neural Network and CART model**   
**Steps involved should be:**

**a) Data Import (Target variable is "Attrition" column)**  
**b) Split the data in Dev & Hold Out sample (70:30)**  
**c) Perform Exploratory Data Analysis**  
**d) Identify columns which are of no use. drop those columns**  
**e) Write Hypothesis and validate the Hypothesis**  
**f) Build Neural Network Model (Development sample)**  
**g) Validate NN model on Hold Out. If need be improvise**  
**h) Build CART Model**  
**i) Validate CART Model**  
**j) Compare NN with CART**  
**k) Combine NN and CART into Ensemble Model**  
**l) Check whether Ensemble Model Performance outperforms the individual CART & NN model**

**By**

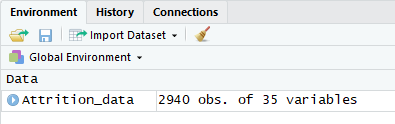
Akshay Surve

**Answer**

**1] Importing Dataset**

Attrition\_data <- read.csv("C:/Users/DELL/Desktop/Akshay/Group Assignments/Group Assignment 4 DM/HR\_Attrition.csv", header=TRUE)

str(Attrition\_data)



summary(Attrition\_data)

Age Attrition BusinessTravel DailyRate

Min. :18.00 No :2466 Non-Travel : 300 Min. : 102.0

1st Qu.:30.00 Yes: 474 Travel\_Frequently: 554 1st Qu.: 465.0

Median :36.00 Travel\_Rarely :2086 Median : 802.0

Mean :36.92 Mean : 802.5

3rd Qu.:43.00 3rd Qu.:1157.0

Max. :60.00 Max. :1499.0

Department DistanceFromHome Education EducationField

Human Resources : 126 Min. : 1.000 Min. :1.000 Human Resources : 54

Research & Development:1922 1st Qu.: 2.000 1st Qu.:2.000 Life Sciences :1212

Sales : 892 Median : 7.000 Median :3.000 Marketing : 318

Mean : 9.193 Mean :2.913 Medical : 928

3rd Qu.:14.000 3rd Qu.:4.000 Other : 164

Max. :29.000 Max. :5.000 Technical Degree: 264

EmployeeCount EmployeeNumber EnvironmentSatisfaction Gender HourlyRate

Min. :1 Min. : 1.0 Min. :1.000 Female:1176 Min. : 30.00

1st Qu.:1 1st Qu.: 735.8 1st Qu.:2.000 Male :1764 1st Qu.: 48.00

Median :1 Median :1470.5 Median :3.000 Median : 66.00

Mean :1 Mean :1470.5 Mean :2.722 Mean : 65.89

3rd Qu.:1 3rd Qu.:2205.2 3rd Qu.:4.000 3rd Qu.: 84.00

Max. :1 Max. :2940.0 Max. :4.000 Max. :100.00

JobInvolvement JobLevel JobRole JobSatisfaction

Min. :1.00 Min. :1.000 Sales Executive :652 Min. :1.000

1st Qu.:2.00 1st Qu.:1.000 Research Scientist :584 1st Qu.:2.000

Median :3.00 Median :2.000 Laboratory Technician :518 Median :3.000

Mean :2.73 Mean :2.064 Manufacturing Director :290 Mean :2.729

3rd Qu.:3.00 3rd Qu.:3.000 Healthcare Representative:262 3rd Qu.:4.000

Max. :4.00 Max. :5.000 Manager :204 Max. :4.000

(Other) :430

MaritalStatus MonthlyIncome MonthlyRate NumCompaniesWorked Over18 OverTime

Divorced: 654 Min. : 1009 Min. : 2094 Min. :0.000 Y:2940 No :2108

Married :1346 1st Qu.: 2911 1st Qu.: 8045 1st Qu.:1.000 Yes: 832

Single : 940 Median : 4919 Median :14236 Median :2.000

Mean : 6503 Mean :14313 Mean :2.693

3rd Qu.: 8380 3rd Qu.:20462 3rd Qu.:4.000

Max. :19999 Max. :26999 Max. :9.000

PercentSalaryHike PerformanceRating RelationshipSatisfaction StandardHours StockOptionLevel

Min. :11.00 Min. :3.000 Min. :1.000 Min. :80 Min. :0.0000

1st Qu.:12.00 1st Qu.:3.000 1st Qu.:2.000 1st Qu.:80 1st Qu.:0.0000

Median :14.00 Median :3.000 Median :3.000 Median :80 Median :1.0000

Mean :15.21 Mean :3.154 Mean :2.712 Mean :80 Mean :0.7939

3rd Qu.:18.00 3rd Qu.:3.000 3rd Qu.:4.000 3rd Qu.:80 3rd Qu.:1.0000

Max. :25.00 Max. :4.000 Max. :4.000 Max. :80 Max. :3.0000

TotalWorkingYears TrainingTimesLastYear WorkLifeBalance YearsAtCompany YearsInCurrentRole

Min. : 0.00 Min. :0.000 Min. :1.000 Min. : 0.000 Min. : 0.000

1st Qu.: 6.00 1st Qu.:2.000 1st Qu.:2.000 1st Qu.: 3.000 1st Qu.: 2.000

Median :10.00 Median :3.000 Median :3.000 Median : 5.000 Median : 3.000

Mean :11.28 Mean :2.799 Mean :2.761 Mean : 7.008 Mean : 4.229

3rd Qu.:15.00 3rd Qu.:3.000 3rd Qu.:3.000 3rd Qu.: 9.000 3rd Qu.: 7.000

Max. :40.00 Max. :6.000 Max. :4.000 Max. :40.000 Max. :18.000

YearsSinceLastPromotion YearsWithCurrManager

Min. : 0.000 Min. : 0.000

1st Qu.: 0.000 1st Qu.: 2.000

Median : 1.000 Median : 3.000

Mean : 2.188 Mean : 4.123

3rd Qu.: 3.000 3rd Qu.: 7.000

Max. :15.000 Max. :17.000

**2] Identifying Unnecessary Variables (Columns) Which Are Irrelevant And Removing Them.**

Dropping the variables (columns) having no variability.

Therefore, Drop **Over18** as there is no variability, all are Y.

Drop **Employee\_Count** as there is no variability, all are 1.

Drop **Standard\_Hours** as there is no variability, all are 80.

Drop **Employee\_Number** as it is just an identifier.

Attrition\_data$Over18 <- NULL

Attrition\_data$EmployeeCount <- NULL

Attrition\_data$StandardHours <- NULL

Attrition\_data$EmployeeNumber <- NULL

Attrition\_data = Attrition\_data[-c(9,10,22,27)]

**3] Performing Exploratory Data Analysis**

summary(Attrition\_data$Attrition)

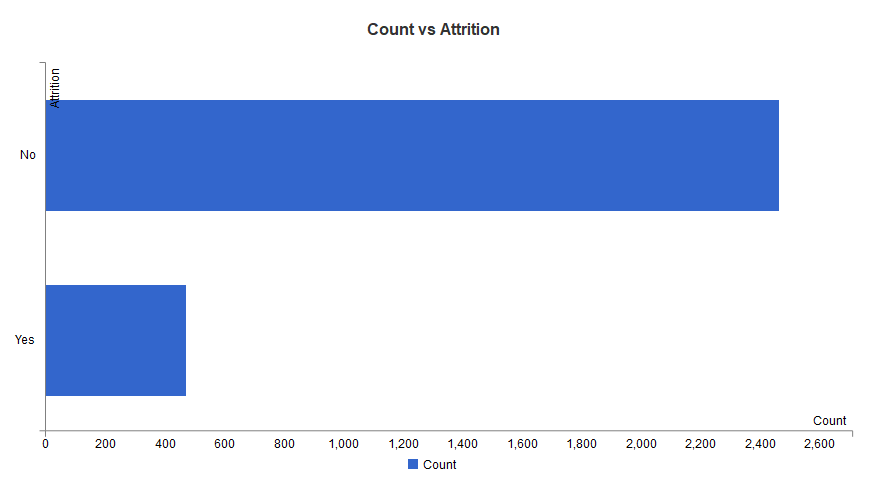
No Yes

2466 474

prop.table(summary(Attrition\_data$Attrition))

No Yes

0.8387755 0.1612245



**Total attrition rate among the employees is (474/2466) = 16.122%**

Now,

Running CART model on whole data to find major factors influencing attrition rate of employee.

The ratio of people in the company to people leaving the company is 1:5.

library(rpart)

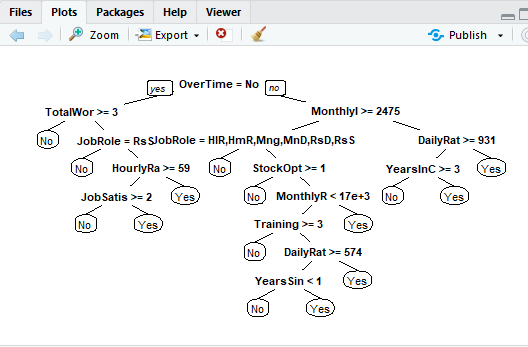
library(rpart.plot)

CART\_model = rpart(Attrition ~., data=Attrition\_data, method="class")

plot(CART\_model)

text(CART\_model, digits = 4, cex = 0.6)

prp(CART\_model,cex = 0.6)



From CART MODEL it can be seen that **Total\_Working\_Years, Overtime, Monthly\_Income, Job\_Role** are the **most influencing factors on attrition** of employee.

apply rpivottable to original dataset,

library(rpivotTable)

rpivotTable(Attrition\_data)

Comparing these most influencing factors with attrition rate of employees

* **Total working years**

table(Attrition\_data$TotalWorkingYears)

prop.table(table(Attrition\_data$TotalWorkingYears))

0 1 2 3 4 5 6

0.0074829932 0.0551020408 0.0210884354 0.0285714286 0.0428571429 0.0598639456 0.0850340136

7 8 9 10 11 12 13

0.0551020408 0.0700680272 0.0653061224 0.1374149660 0.0244897959 0.0326530612 0.0244897959

14 15 16 17 18 19 20

0.0210884354 0.0272108844 0.0251700680 0.0224489796 0.0183673469 0.0149659864 0.0204081633

21 22 23 24 25 26 27

0.0231292517 0.0142857143 0.0149659864 0.0122448980 0.0095238095 0.0095238095 0.0047619048

28 29 30 31 32 33 34

0.0095238095 0.0068027211 0.0047619048 0.0061224490 0.0061224490 0.0047619048 0.0034013605

35 36 37 38 40

0.0020408163 0.0040816327 0.0027210884 0.0006802721 0.0013605442

table(Attrition\_data$TotalWorkingYears, Attrition\_data$Attrition)

No Yes

0 12 10

1 82 80

2 44 18

3 66 18

4 102 24

5 144 32

6 206 44

7 126 36

8 174 32

9 172 20

10 354 50

11 58 14

12 86 10

13 66 6

14 54 8

15 70 10

16 68 6

17 60 6

18 46 8

19 38 6

20 56 4

21 66 2

22 38 4

23 40 4

24 30 6

25 26 2

26 26 2

27 14 0

28 26 2

29 20 0

30 14 0

31 16 2

32 18 0

33 12 2

34 8 2

35 6 0

36 12 0

37 8 0

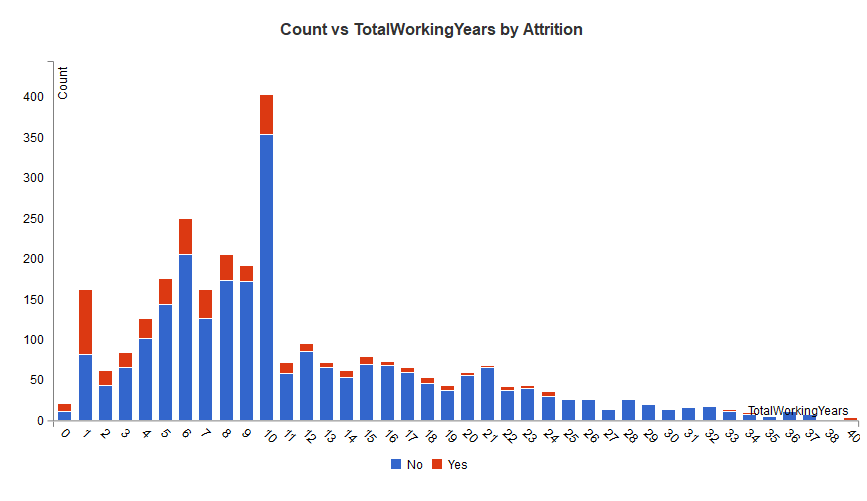
38 2 0

40 0 4

summary(Attrition\_data$TotalWorkingYears)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.00 6.00 10.00 11.28 15.00 40.00



From bar chart it can be seen that Attrition rate is high among people with low work experience and it decreases drastically as number of working year increases.

People with less Experience are leaving the job more.

* **Overtime**

table(Attrition\_data$OverTime)

No Yes

2108 832

prop.table(table(Attrition\_data$OverTime))

No Yes

0.7170068 0.2829932

table(Attrition\_data$OverTime, Attrition\_data$Attrition)

No Yes

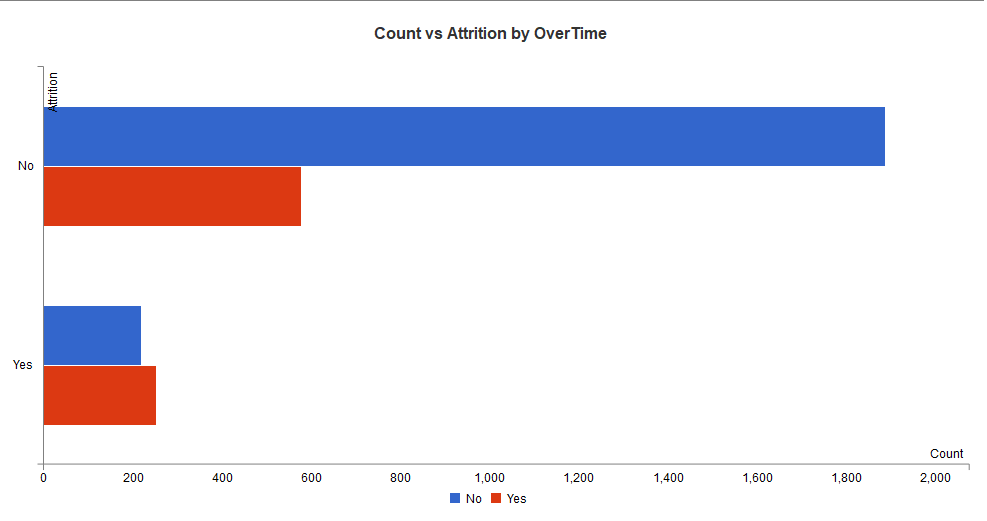
No 1888 220

Yes 578 254

summary(Attrition\_data$OverTime)

No Yes

2108 832



The attrition rate in people doing overtime is (832/2940) = 0.2829 I.e 28.29% which is quite high. And Attrition rate is higher among employees doing overtime.

* **Monthly income**

table(Attrition\_data$MonthlyIncome)

prop.table(table(Attrition\_data$MonthlyIncome))

table(Attrition\_data$MonthlyIncome, Attrition\_data$Attrition)

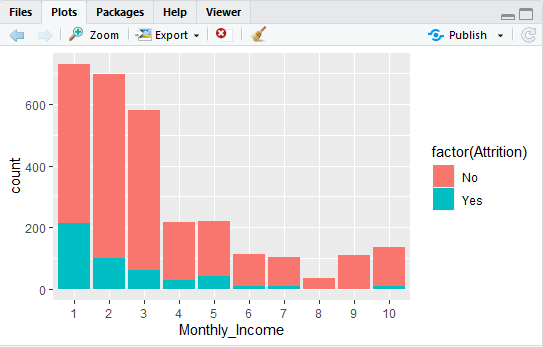
summary(Attrition\_data$MonthlyIncome)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1009 2911 4919 6503 8380 19999

Monthly\_Income <- cut(Attrition\_data$MonthlyIncome, 10, include.lowest = TRUE, labels=c(1,2,3,4,5,6,7,8,9,10))

ggplot(Attrition\_data, aes(Monthly\_Income, ..count.., fill = factor(Attrition))) + geom\_bar(position="stack")



Attrition rate is higher among the people with low salary. It decreases with higher salaries.

* **Job role**

table(Attrition\_data$JobRole)

Healthcare Representative Human Resources Laboratory Technician

262 104 518

Manager Manufacturing Director Research Director

204 290 160

Research Scientist Sales Executive Sales Representative

584 652 166

prop.table(table(Attrition\_data$JobRole))

Healthcare Representative Human Resources Laboratory Technician

0.08911565 0.03537415 0.17619048

Manager Manufacturing Director Research Director

0.06938776 0.09863946 0.05442177

Research Scientist Sales Executive Sales Representative

0.19863946 0.22176871 0.05646259

table(Attrition\_data$JobRole, Attrition\_data$Attrition)

No Yes

Healthcare Representative 244 18

Human Resources 80 24

Laboratory Technician 394 124

Manager 194 10

Manufacturing Director 270 20

Research Director 156 4

Research Scientist 490 94

Sales Executive 538 114

Sales Representative 100 66

summary(Attrition\_data$JobRole)

Healthcare Representative Human Resources Laboratory Technician

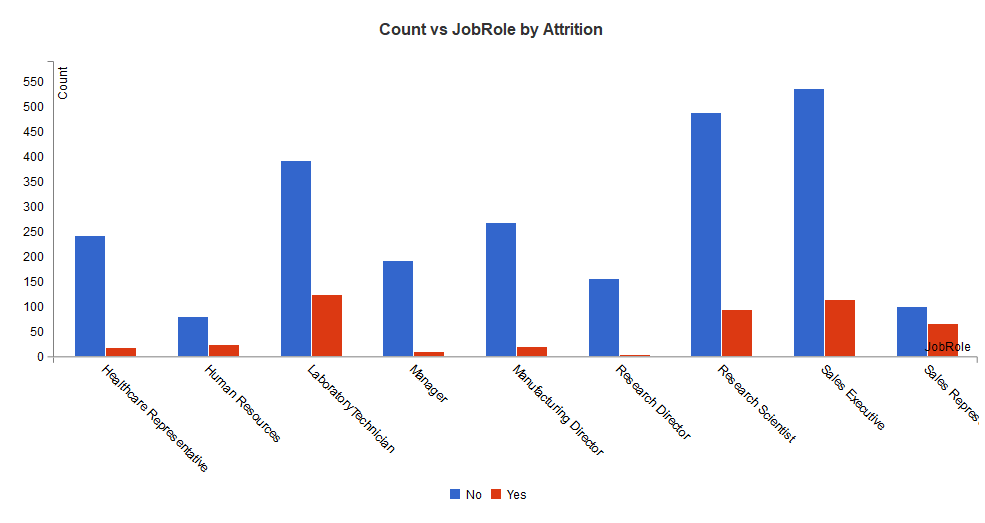
262 104 518

Manager Manufacturing Director Research Director

204 290 160

Research Scientist Sales Executive Sales Representative

584 652 166



Attrition rate is higher among the group of people working as Laboratory technicians, Research scientist, Sales executive and Sales representative.

**Hypothesis**

1.     Overtime

**H0 = Overtime Plays a Significant factor for attrition.**

**H1= Overtime is not a Significant factor for attrition.**

2.     Job Role

**H0 = Managers and Directors are significantly less in the attrition, they retain in the company**

**H1= Managers and Directors are not significantly less in the attrition, they retain in the company**

3.     Experience

**H0 = Attrition are more where the employee having less years of experience in the Company**

**H1= Attrition are not more where the employee having less years of experience in the Company**

4.     Ensemble Model

**H0 = Ensembling with different Technique will have better prediction**

**H1= Ensembling with different Technique will not have better prediction**

Rearranging data for checking correlation among variables

It is necessary for correlation plot as it accepts only numeric values

Attrition\_data$Attrition <- as.numeric(ifelse(Attrition\_data$Attrition =="No",0,1))

Attrition\_data$BusinessTravel <- as.numeric(factor(Attrition\_data$BusinessTravel,

levels=c('Non-Travel','Travel\_Rarely','Travel\_Frequently'),

labels = c(1,2,3)))

Attrition\_data$Department <- as.numeric(factor(Attrition\_data$Department,

levels=c('Sales','Research & Development','Human Resources'),

labels = c(1,2,3)))

Attrition\_data$EducationField <- as.numeric(factor(Attrition\_data$EducationField,

levels=c('Life Sciences','Medical','Marketing','Technical Degree', 'Human Resources','Other'),

labels = c(1,2,3,4,5,6)))

Attrition\_data$Gender <- as.numeric(factor(Attrition\_data$Gender,

levels=c('Male', 'Female'),

labels = c(1,2)))

Attrition\_data$JobRole <- as.numeric(factor(Attrition\_data$JobRole,

levels=c('Healthcare Representative', 'Human Resources','Laboratory Technician','Manager','Research Scientist','Sales Executive','Sales Representative','Manufacturing Director','Research Director'),

labels = c(1,2,3,4,5,6,7,8,9)))

Attrition\_data$MaritalStatus <- as.numeric(factor(Attrition\_data$MaritalStatus,

levels=c('Single','Married','Divorced'),

labels = c(1,2,3)))

Attrition\_data$OverTime <- as.numeric(factor(Attrition\_data$OverTime,

levels=c('Yes','No'), labels = c(1,2)))

Creating Correlation Plot

library(corrplot)

cor(Attrition\_data)

corrplot(cor(Attrition\_data))

Correlation plot further gives idea about which variables to consider, i.e neglecting the variables having high correlation with Target variable (in this case Attrition).

From the Correlation plot, we have dropped the following variables:

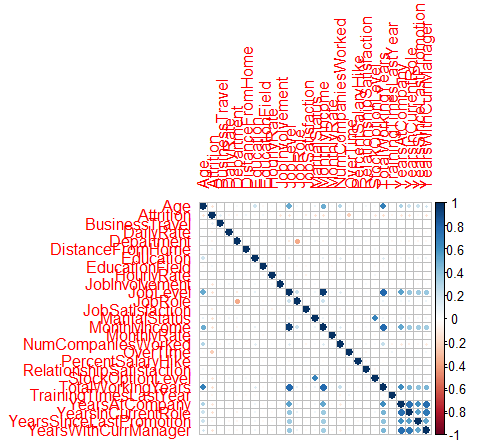
a) YearsAtCompany

b) YearsInCurrentRole

c) YearsSinceLastPromotion

d) Monthly Income

e) Performance Rating



**4] Split the data in Dev & Hold Out sample (70:30).**

set.seed(555)

library(caTools)

n=nrow(Attrition\_data)

split=sample(c(TRUE,FALSE),n,replace=TRUE,prob = c(0.70,0.30))

Creating dev and holdout sample sets

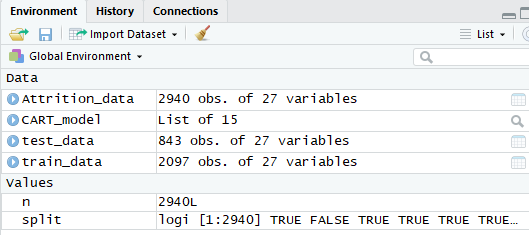
train\_data = Attrition\_data[split,]

test\_data = Attrition\_data[!split,]

The imported data after dropping of redundant and correlated variables has been split into training and test sets using sample. Split on Attrition variable with split ratio=0.7 from “caTools” package.

All the models have been trained on training set and validated on test set.

The Target variable has been encoded as Yes-1 No-0.



**5] Building Neural Network Model on Development sample**

Dataset been reordered as-Attrition followed by all numeric fields and categorical fields for easier cleaning of data for model building.

All the categorical variables have been encoded as factors and all the variables are scaled to ensure easier and faster computations

nn.train.us <- train\_data

nn.test.us <- test\_data

set.seed(555)

library(neuralnet)

library(prediction)

dim(nn.train.us)

[1] 2097 27

nn\_us<-neuralnet(Attrition ~ Age + BusinessTravel + DailyRate + Department+

DistanceFromHome +Education +EducationField +

HourlyRate + JobInvolvement +JobLevel +JobRole +JobSatisfaction +

MaritalStatus + MonthlyRate + NumCompaniesWorked +OverTime +

PercentSalaryHike +RelationshipSatisfaction+StockOptionLevel+TotalWorkingYears+

TrainingTimesLastYear+ YearsAtCompany + YearsInCurrentRole +

YearsSinceLastPromotion + YearsWithCurrManager,

data = nn.train.us,

hidden = 4,err.fct = "sse",

linear.output = FALSE,

lifesign = "full",

lifesign.step = 10,

stepmax = 2000,

learningrate = 1,

threshold = 0.1)

Attrition (Dependent variable) is compared to all Independent Variables .

Parameters considered:

4 hidden layers.

Threshold set to 0.1

Stepmax: 2000.

predictionNN<-predict(nn\_us,nn.test.us,type=("class"))

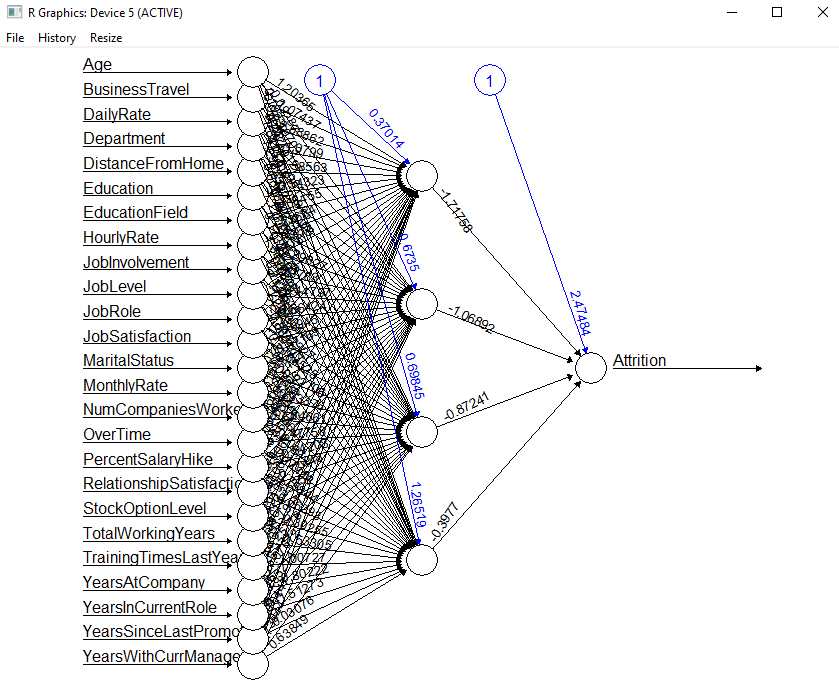
table(predictionNN)

predictionNN

0.170545081580069 0.170545081580677

841 2

plot(nn\_us)



summary(nn\_us)

Length Class Mode

call 11 -none- call

response 2097 -none- numeric

covariate 52425 -none- numeric

model.list 2 -none- list

err.fct 1 -none- function

act.fct 1 -none- function

linear.output 1 -none- logical

data 27 data.frame list

exclude 0 -none- NULL

net.result 1 -none- list

weights 1 -none- list

generalized.weights 1 -none- list

startweights 1 -none- list

result.matrix 112 -none- numeric

Confussion Matrix

t3 <- table(nn.test.us$Attrition, predictionNN)

predictionNN

0 1

0 725 2

1 116 39

Neural Network model accuracy

(t3[1]+t3[4])/(nrow(nn.test.us))

**[1] 0.8600237**

Total accuracy of the neural network after testing it on test data is 86.0023% which is significantly high.

**Summary of Neural Network model: From Confusion Matrix**

|  |  |
| --- | --- |
| **Accuracy=(TP+TN)/total** | **0.8662** |
| **Misclassification Rate or error rate** | **0.13378** |
| **Sensitivity or recall or True Positive Rate** | **0.94895** |
| **Specificity = TN/actual no** | **0.95121** |
|  |  |

**6] Building CART Model on train data**

library(rpart)

library(rpart.plot)

cart = rpart.control(minsplit=100, minbucket = 10, cp = 0, xval = 10)

ptree <- rpart(formula = train\_data$Attrition ~ .,

data = train\_data[,2:26], method = "class",

control =cart)

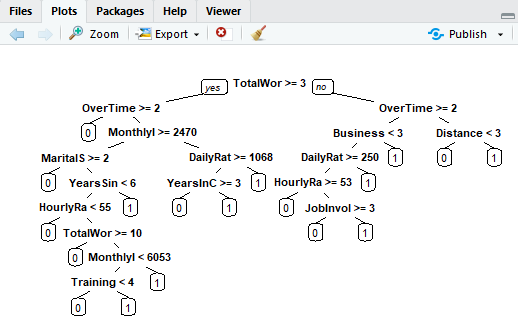
rpart.plot(ptree)

CART\_train = rpart(Attrition ~., data=train\_data, method="class")

plot(CART\_train)

text(CART\_train, digits = 4, cex = 0.6)

prp(CART\_train,cex = 0.6)



print(CART\_train)

n= 2097

node), split, n, loss, yval, (yprob)

\* denotes terminal node

1) root 2097 358 No (0.82927992 0.17072008)

2) TotalWorkingYears>=2.5 1920 272 No (0.85833333 0.14166667)

4) OverTime=No 1378 117 No (0.91509434 0.08490566) \*

5) OverTime=Yes 542 155 No (0.71402214 0.28597786)

10) MonthlyIncome>=2469.5 472 104 No (0.77966102 0.22033898)

20) JobRole=Healthcare Representative,Human Resources,Manager,Manufacturing Director,Research Director,Research Scientist 260 30 No (0.88461538 0.11538462) \*

21) JobRole=Laboratory Technician,Sales Executive,Sales Representative 212 74 No (0.65094340 0.34905660)

42) StockOptionLevel>=0.5 126 27 No (0.78571429 0.21428571)

84) MonthlyIncome>=3860 97 12 No (0.87628866 0.12371134) \*

85) MonthlyIncome< 3860 29 14 Yes (0.48275862 0.51724138)

170) MonthlyIncome< 2759.5 12 2 No (0.83333333 0.16666667) \*

171) MonthlyIncome>=2759.5 17 4 Yes (0.23529412 0.76470588) \*

43) StockOptionLevel< 0.5 86 39 Yes (0.45348837 0.54651163)

86) MonthlyIncome< 8163.5 68 29 No (0.57352941 0.42647059)

172) MonthlyRate< 17126 46 13 No (0.71739130 0.28260870)

344) TrainingTimesLastYear>=2.5 21 1 No (0.952380 0.0476) \*

345) TrainingTimesLastYear< 2.5 25 12 No (0.52000000 0.4800)

690) DailyRate>=952.5 9 0 No (1.00000000 0.00000000) \*

691) DailyRate< 952.5 16 4 Yes (0.25000000 0.75000000) \*

173) MonthlyRate>=17126 22 6 Yes (0.27272727 0.72727273) \*

87) MonthlyIncome>=8163.5 18 0 Yes (0.00000000 1.00000000) \*

11) MonthlyIncome< 2469.5 70 19 Yes (0.27142857 0.72857143)

22) DailyRate>=1067.5 22 9 No (0.59090909 0.40909091)

44) YearsInCurrentRole>=2.5 10 0 No (1.00000000 0.00000000) \*

45) YearsInCurrentRole< 2.5 12 3 Yes (0.25000000 0.75000000) \*

23) DailyRate< 1067.5 48 6 Yes (0.12500000 0.87500000) \*

3) TotalWorkingYears< 2.5 177 86 No (0.51412429 0.48587571)

6) OverTime=No 128 50 No (0.60937500 0.39062500)

12) BusinessTravel=Non-Travel,Travel\_Rarely 112 36 No (0.67857 0.321)

24) DailyRate>=249.5 99 25 No (0.74747475 0.25252525)

48) HourlyRate>=53 68 9 No (0.86764706 0.13235294) \*

49) HourlyRate< 53 31 15 Yes (0.48387097 0.51612903)

98) JobInvolvement>=2.5 19 5 No (0.73684211 0.26315789) \*

99) JobInvolvement< 2.5 12 1 Yes (0.08333333 0.91666667) \*

25) DailyRate< 249.5 13 2 Yes (0.15384615 0.84615385) \*

13) BusinessTravel=Travel\_Frequently 16 2 Yes (0.12500000 0.87500)\*

7) OverTime=Yes 49 13 Yes (0.26530612 0.73469388)

14) DistanceFromHome< 2.5 10 3 No (0.70000000 0.30000000) \*

15) DistanceFromHome>=2.5 39 6 Yes (0.15384615 0.84615385) \*

Predict the test data

Predict\_CART <- predict(CART\_train, newdata=test\_data, type="class")

summary(Predict\_CART)

No Yes

768 75

Confusion matrix

cm <- table(test\_data$Attrition, Predict\_CART)

print(cm)

Predict\_CART

No Yes

No 701 26

Yes 67 49

CART model accuracy

(t1[1]+t1[4])/(nrow(test\_data))

**[1] 0.8896797**

The CART model after testing on test data fives an accuracy of 88.9679% which is slightly higher than Neural network model. But there is no drastic increase in the accuracy.

**Summary of CART model: From Confusion Matrix**

|  |  |
| --- | --- |
| **Accuracy=(TP+TN)/total** | **0.8896** |
| **Misclassification Rate or error rate** | **0.11032** |
| **Sensitivity or recall or True Positive Rate** | **0.934667** |
| **Specificity = TN/actual no** | **0.65333** |
|  |  |

**7]** **Comparing Neural Network with CART**

|  |  |  |
| --- | --- | --- |
|  | Neural Network | CART model |
| Accuracy=(TP+TN)/total | 0.8662 | 0.8896 |
| Misclassification Rate or error rate | 0.13378 | 0.11032 |
| Sensitivity or recall or True Positive Rate | 0.94895 | 0.934667 |
| Specificity = TN/actual no | 0.95121 | 0.65333 |

From table it can be seen that CART model gives slightly better accuracy than Neural Network model.

**8]** **Combine NN and CART into Ensemble Model**

**predictions <- data.frame(predictionCart= Predict\_CART,predictionNN = predictionNN)**

**predictions$predictionEnsemble <- as.factor(ifelse(predictions$predictionCart=='Yes'**

**& predictions$predictionNN=='Yes','Yes','No'))**

**Confusion Matrix**

**t4 <- table(test\_data$Attrition, predictions$predictionEnsemble)**

No Yes

No 732 8

Yes 59 83

**Ensembeling Accuracy**

**(t4[1]+t4[4])/(nrow(test\_data))**

**[1] 0.92403628**

**Ensemble model gives much higher accuracy than the individual accuracies of CART and Neural Network model.**

**9] Check whether Ensemble Model Performance outperforms the individual CART & NN model**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Neural Network | CART model | Ensemble model |
| Accuracy=(TP+TN)/total | 0.8662 | 0.8896 | 0.92402 |
| Misclassification Rate or error rate | 0.13378 | 0.11032 | 0.07596 |
| Sensitivity or recall or True Positive Rate | 0.94895 | 0.934667 | 0.92541 |
| Specificity = TN/actual no | 0.95121 | 0.65333 | 0.91208 |

**From above it can be clearly seen that ensemble model gives higher accuracy than that of individual model accuracies.**

**From this we can say that ensemble model outperforms the individual Neural Network and CART model.**

**10] Hypothesis Validation**

**Conclusion & Solutions to Hypothesis:**

As per the analysis using the CART, NN & Ensembling Model, we conclude that all Null Hypothesis are accepted all alternative are rejected.

Ensembling Model is giving better accuracy than CART & NN

1. Overtime

**H0 = Overtime Plays a Significant factor for attrition.**

**H1= Reject alternate**

2. Job Role

**H0 = Managers and Directors are significantly less in the attrition, they retain in the company**

**H1= Reject alternate**

3. Experience

**H0 = Attrition are more where the employee having less years of experience in the Company**

**H1= Reject alternate**

4. Ensemble Model

**H0 = Ensembling with different Technique will have better prediction**

**H1= Reject alternate**