Assignment 17.1

Name : Y Vasudev

Batch : DA with R , Excel and Tableau

1. Use the below given data set

Data Set

2. Perform the below given activities:

a. Create classification model using logistic regression model

b. verify model goodness of fit

c. Report the accuracy measures

d. Report the variable importance

e. Report the unimportant variables

f. Interpret the results

g. Visualize the results

#a. Create classification model using logistic regression model

#using dataset cs2m

#reading the dataset

cs2m <- read.csv("D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\17 ENSEMBLE MODELS\\cs2m.csv")

View(cs2m)

#logistic regression

model<- glm(classe~cvtd\_timestamp+total\_accel\_belt+yaw\_dumbbell+roll\_forearm+accel\_forearm\_y, data = cs2m ,family= binomial(link='logit'))

model

summary(model)

#classification

library(caTools)

library(tree)

#splitting

set.seed(1)

split<- sample.split(cs2m$classe,SplitRatio = 0.70)

cs2mTrain <- subset(cs2m,split == TRUE)

cs2mTest<- subset(cs2m, split == FALSE)

modelClassTree<- tree(classe~cvtd\_timestamp+total\_accel\_belt+yaw\_dumbbell+roll\_forearm+accel\_forearm\_y,data = cs2mTrain)

plot(modelClassTree)

text(modelClassTree,pretty = 0 ,cex=0.75)

pred<- predict(modelClassTree,newdata= cs2mTest)

predict<- predict(model,type="response")

head(predict,3)

cs2m$predict <- predict

cs2m$predictROUND<- round(predict,digits = 0)

#confusion matrix

table(cs2m$classe,predict>= 0.5)

sum<- sum(table(cs2m$classe,predict>= 0.5))

#b. verify model goodness of fit

#c. Report the accuracy measures

#f. Interpret the results

#Answer for b & c & f

#interpretation, Accuracy and model goodness of our model

summary(model)

#accuracy of our model

accuracy<- (1185+679)/(2266)

accuracy

#0.8225949

library(verification)

library(AUC)

predictTrain<- predict(model,cs2m,type="response")

table(cs2m$classe,predictTrain >=0.5)

head(predictTrain,3)

auc(cs2m$classe,predictTrain)

#model goodness

#\*\*\*\*NOTE\*\*\*\*

#Area under the curve: 0.9333333

#also our AIC is less which is measure of good model

#NULL deviance is also less which is good for model

#Residual deviance is also less model

#by this all things we conclude that our model is good and fit

#e. Report the unimportant variables

library(MASS)

step\_fit<- stepAIC(model,method ="backward")

summary(step\_fit)

confint(step\_fit)

#thus by this method we get our best model and variable cvtd\_timestamp is not as much important y this method

#some test

#ANOVA on base model

anova(model,test = 'Chisq')

#ANOVA from reduced model after applying the Step AIC

anova(step\_fit,test = 'Chisq')

#check for multicollinearity

library(car)

vif(model)

vif(step\_fit)

#d. Report the variable importance

library(caret)

varImp(step\_fit)

#g. Visualize the results

#plot the fitted model

plot(model$fitted.values)

#plot glm

library(ggplot2)

ggplot(cs2mTrain, aes(x=yaw\_dumbbell, y=classe)) + geom\_point() +

stat\_smooth(method="glm", family="binomial", se=FALSE)