#col\_names

"name" "MDVP.Fo.Hz." "MDVP.Fhi.Hz." "MDVP.Flo.Hz."

"MDVP.Jitter..." "MDVP.Jitter.Abs." "MDVP.RAP" "MDVP.PPQ"

"Jitter.DDP" "MDVP.Shimmer" "MDVP.Shimmer.dB." "Shimmer.APQ3"

"Shimmer.APQ5" "MDVP.APQ" "Shimmer.DDA" "NHR"

"HNR" "status" "RPDE" "DFA"

"spread1" "spread2" "D2" "PPE"

------------------------------------------

#col\_names\_of\_m

"name" "MDVP.Fo.Hz." "MDVP.Fhi.Hz." "MDVP.Flo.Hz."

[5] "MDVP.Jitter..." "MDVP.Jitter.Abs." "MDVP.RAP" "MDVP.PPQ"

[9] "Jitter.DDP" "MDVP.Shimmer" "MDVP.Shimmer.dB." "Shimmer.APQ3"

[13] "Shimmer.APQ5" "MDVP.APQ" "Shimmer.DDA" "NHR"

[17] "HNR" "status" "RPDE" "DFA"

[21] "spread1" "spread2" "D2" "PPE"

#read\_table

m<-read.table("G:\\voice R\\main2.csv",header=TRUE,sep=",")

m

#summary

summary(m)

#names\_of\_cols

names(m)

#missing\_values

is.na(m)

#whole\_boxplot

png(file="boxplot.png")

boxplot(m,las=2,names=c("name","MDVP.Fo.Hz.","MDVP.Fhi.Hz.","MDVP.Flo.Hz.","MDVP.Jitter...","MDVP.Jitter.Abs.","MDVP.RAP","MDVP.PPQ","Jitter.DDP","MDVP.Shimmer","MDVP.Shimmer.dB.","Shimmer.APQ3","Shimmer.APQ5","MDVP.APQ","Shimmer.DDA","NHR","HNR","status","RPDE","DFA","spread1","spread2","D2","PPE"))

dev.off()

#outlier\_detection\_single\_col

outlier\_values<- boxplot.stats(m$MDVP.Fhi.Hz)$out

boxplot(m$MDVP.Fhi.Hz, main="MDVP.Fhi.Hz", boxwex=0.1)

mtext(paste("Outliers: ", paste(outlier\_values, collapse=", ")), cex=0.6)

#outliers\_treatment

str(m)

summary(m)

m\_sub<-subset(m,MDVP.Fo.Hz. <210 &MDVP.Fhi.Hz. <300 &MDVP.Flo.Hz. <160 )

summary(m\_sub)

str(m\_sub)

png(file="boxplot.png")

boxplot(m\_sub,las=2,names=c("name","MDVP.Fo.Hz.","MDVP.Fhi.Hz.","MDVP.Flo.Hz.","MDVP.Jitter...","MDVP.Jitter.Abs.","MDVP.RAP","MDVP.PPQ","Jitter.DDP","MDVP.Shimmer","MDVP.Shimmer.dB.","Shimmer.APQ3","Shimmer.APQ5","MDVP.APQ","Shimmer.DDA","NHR","HNR","status","RPDE","DFA","spread1","spread2","D2","PPE"))

dev.off()

#neural\_network

library(MASS)

library(neuralnet)

str(m\_sub)

m\_sub$MDVP.Fo.Hz.<-(m\_sub$MDVP.Fo.Hz.-min(m\_sub$MDVP.Fo.Hz.)/max(m\_sub$MDVP.Fo.Hz.)-min(m\_sub$MDVP.Fo.Hz.))

m\_sub$MDVP.Fhi.Hz.<-(m\_sub$MDVP.Fhi.Hz.-min(m\_sub$MDVP.Fhi.Hz.)/max(m\_sub$MDVP.Fhi.Hz.)-min(m\_sub$MDVP.Fhi.Hz.))

m\_sub$MDVP.Flo.Hz.<-(m\_sub$MDVP.Flo.Hz.-min(m\_sub$MDVP.Flo.Hz.)/max(m\_sub$MDVP.Flo.Hz.)-min(m\_sub$MDVP.Flo.Hz.))

m\_sub$HNR<-(m\_sub$HNR-min(m\_sub$HNR)/max(m\_sub$HNR)-min(m\_sub$HNR))

m\_sub$spread1<-(m\_sub$spread1-min(m\_sub$spread1)/max(m\_sub$spread1)-min(m\_sub$spread1))

m\_sub$PPE<-(m\_sub$PPE-min(m\_sub$PPE)/max(m\_sub$PPE)-min(m\_sub$PPE))

#data\_partitioning

ind<-sample(2,nrow(m\_sub),replace=TRUE,prob=c(0.7,0.3))

training<-m\_sub[ind==1,]

test<-m\_sub[ind==2,]

#nn

library(neuralnet)

n<-neuralnet(status~MDVP.Fo.Hz.+MDVP.Fhi.Hz.+MDVP.Flo.Hz.+MDVP.Jitter...+MDVP.Jitter.Abs.+MDVP.RAP+MDVP.PPQ+Jitter.DDP+MDVP.Shimmer+MDVP.Shimmer.dB.+Shimmer.APQ3+Shimmer.APQ5+MDVP.APQ+Shimmer.DDA+NHR+HNR+RPDE+DFA+spread1+spread2+D2+PPE,data=training,hidden=3,err.fct="ce",linear.output=FALSE)

plot(n)

#miss\_classification(half\_done)

output<-compute(n,training[,1])

p1<-output$net.result

pred1<-ifelse(p1>0.5,1,0)

tab1<-table(pred1,training$status)

outlierKD<- function(v,MDVP.Flo.Hz.) {

k <- eval(substitute(k),eval(v))

na1 <- sum(is.na(var\_name))

m1 <- mean(k, na.rm = T)

par(mfrow=c(2, 2), oma=c(0,0,3,0))

boxplot(k, main="With outliers")

hist(var\_name, main="With outliers", xlab=NA, ylab=NA)

outlier<- boxplot.stats(k)$out

mo<- mean(outlier)

k <- ifelse(k %in% outlier, NA, k)

boxplot(k, main="Without outliers")

hist(k, main="Without outliers", xlab=NA, ylab=NA)

title("Outlier Check", outer=TRUE)

na2 <- sum(is.na(k))

cat("Outliers identified:", na2 - na1, "n")

cat("Propotion (%) of outliers:", round((na2 - na1) / sum(!is.na(k))\*100, 1), "n")

cat("Mean of the outliers:", round(mo, 2), "n")

m2 <- mean(k, na.rm = T)

cat("Mean without removing outliers:", round(m1, 2), "n")

cat("Mean if we remove outliers:", round(m2, 2), "n")

response<- readline(prompt="Do you want to remove outliers and to replace with NA? [yes/no]: ")

if(response == "y" | response == "yes"){

dt[as.character(substitute(MDVP.Flo.Hz.))] <- invisible(k)

assign(as.character(as.list(match.call())$dt), dt, envir = .GlobalEnv)

cat("Outliers successfully removed", "n")

return(invisible(dt))

} else{

cat("Nothing changed", "n")

return(invisible(k))

}

}

remove\_outliers<- function(v, na.rm = TRUE, ...) {

qnt<- quantile(v, probs=c(.25, .75), na.rm = na.rm, ...)

H <- 1.5 \* IQR(v, na.rm = na.rm)

y <- v

y[v < (qnt[1] - H)] <- NA

y[v > (qnt[2] + H)] <- NA

y

}

set.seed(1)

v <- rnorm(100)

v <- c(-10, v, 10)

y <- remove\_outliers(v)

png()

par(mfrow = c(1, 2))

boxplot(v)

boxplot(y)

dev.off()

rm.outlier(v, fill = FALSE, median = FALSE, opposite = FALSE)

x[!x %in% boxplot.stats(x)$out]