**Splitting test and training data**

sub<- sample(nrow(mydata),nrow(mydata)\*0.8)  
> train <- mydata[sub,]  
> test <- mydata[-sub,]

**Decision tree**

tree1<- rpart(NRC\_CLASS~L1\_CLASS+L2\_CLASS+S1\_CLASS,data=train,method="class")

tree1 <- rpart(NRC\_CLASS~ SCHOOL\_TYPE+ URBAN\_RURAL+NRC\_CASTE\_CODE,data=train,method="class")

tree1 <- rpart(L1\_CODE~NRC\_MEDIUM+ URBAN\_RURAL+SCHOOL\_TYPE,data=train,method="class")

pred1 <- predict(tree1,test,type="class")

table(train$NRC\_CLASS,predict(tree2,type="class"))

sum(test$NRC\_CLASS==Pred2)/length(Pred2)

confusionMatrix(pred1, test$NRC\_CLASS)

**Clustering : K-means**

fit<- kmeans(newdata,2)

table(mydata$NRC\_CLASS, fit$cluster)

plot(mydata[c("L1\_MARKS","L2\_MARKS")], col=result$cluster)

cluster1 <- final\_result[final\_result$result.cluster==1,]

d<-data.frame(x=sample$L1\_MARKS, y=sample$L2\_MARKS)

plot(d)

mydata<- d

wss<- (nrow(data)-1)\*sum(apply(data,2,var))

for (i in 2:15) wss[i] <- sum(kmeans(data,centers=i)$withinss)

**Silhoutte**cl=kmeans(data[1:100,],5)

>diss=daisy(data[1:100,])

>sk<- silhouette(cl$cl,diss)

>plot(sk)

**Hierarchial clustering**

d <- dist(data[1:100,])

> fit <- hclust(d,method="complete")

>plot(fit)

cdata=data.frame(orgdata$SCHOOL\_TYPE,orgdata$URBAN\_RURAL,orgdata$NRC\_CASTE\_CODE,orgdata$NRC\_MEDIUM,orgdata$L1\_CLASS,orgdata$L2\_CLASS,orgdata$L3\_CLASS,orgdata$S1\_CLASS,orgdata$S2\_CLASS,orgdata$S3\_CLASS,orgdata$NRC\_CLASS,result$cluster)

>head(cdata)