**ALLERGY DATASET 1**

file<-"E:\\Ph.D\\3-JOURNALS\\A2-ALLERGY DSS\\ADSS Codes\\Allergy\_Data1\_ubtrain.csv"

> allergydataset <- read.csv(file, header = FALSE)

> colnames(allergydataset) <- c("housedust","cottondust","aspergilus","pollen", "parthenium", "cockroach","catdander","dosfur","roaddust","oldpdust","PSdust","MilkP","MilkC","curd","coffee","tea","beef","chicken","mutton","egg","fishA","fishB","crab","prawns","shark","avaraikai","banana","beans","beetroot","brinjal","cabbage","capsicum","chillie","cauliflower","carrot","chowchow","corn","cucumber","drumstick","greens","gourds","kovaikai","kothavarai","Lfinger","malli","mango","mushroom","nuckol","onion","peas","potroot","paneer","potato","pumkin","pudina","radish","tomato","tondaikai","vazpooThandu","yams","gram","channa","dhal","maida","oats","ragi","rice","wheat","coconut","oil","garlic","ginger","pepper","tamarind","aginomoto","spices","coco","horlicks","boost","nuts","runningnose","sneeze","cough","wheezeBlocks","headache","itching","swelling","redrashes","FHistory","Sex","Diagnosis")

> table(allergydataset$Diagnosis)

**NORMAL OTHERS RH RH\_O RH\_UT UT UT\_O**

**14 139 242 22 18 258 5**

CONSTANT SAMPLES

> col\_names <- names(allergydataset)

> allergydataset[,col\_names] <- lapply(allergydataset[,col\_names] , factor)

> ideal1<-SmoteClassif(Diagnosis ~ ., allergydataset, C.perc ="balance",dist="HEOM")

> table(ideal$Diagnosis)

**NORMAL OTHERS RH RH\_O RH\_UT UT UT\_O**

**100 100 100 99 99 100 100**

BALANCING ONLY 4 SINGLETON CLASSES

singleton1<-SmoteClassif(Diagnosis ~ ., allergydataset, C.perc =list(NORMAL=1.4,OT=1,RH=1.1,RH\_O=1,RH\_UT=1,UT=0.73,UT\_O=1),dist="HEOM")> table(singleton1$Diagnosis)

**NORMAL OTHERS RH RH\_O RH\_UT UT UT\_O**

**19 139 266 22 18 188 5**

BALANCING ALL 7 CLASSES

full1<-SmoteClassif(Diagnosis ~ ., allergydataset, C.perc =list(NORMAL=0.9,OT=1,RH=0.95,RH\_O=1.3,RH\_UT=1.7,UT=0.75,UT\_O=1),dist="HEOM")

> table(full1$Diagnosis)

**NORMAL OTHERS RH RH\_O RH\_UT UT UT\_O**

**12 139 229 28 30 193 5**

write.csv(full1,"E:\\Ph.D\\3-JOURNALS\\A2-ALLERGY DSS\\ADSS Codes\\Allergy\_Data1\_full.csv")

> write.csv(singleton1,"E:\\Ph.D\\3-JOURNALS\\A2-ALLERGY DSS\\ADSS Codes\\Allergy\_Data1\_single.csv")

> write.csv(ideal1,"E:\\Ph.D\\3-JOURNALS\\A2-ALLERGY DSS\\ADSS Codes\\Allergy\_Data1\_ideal.csv")

> dim(ideal1)

**[1] 698 91**

> dim(singleton1)

**[1] 657 91**

> dim(full1)

**[1] 636 91**

**ALLERGY DATASET 2**

file2<-"E:\\Ph.D\\3-JOURNALS\\A2-ALLERGY DSS\\ADSS Codes\\Allergy\_Data2\_ubtrain.csv"

> allergydataset2 <- read.csv(file2, header = FALSE)

> colnames(allergydataset2) <- c("housedust","cottondust","aspergilus","pollen", "parthenium", "cockroach","catdander","dosfur","roaddust","oldpdust","PSdust","MilkP","MilkC","curd","coffee","tea","beef","chicken","mutton","egg","fishA","fishB","crab","prawns","shark","avaraikai","banana","beans","beetroot","brinjal","cabbage","capsicum","chillie","cauliflower","carrot","chowchow","corn","cucumber","drumstick","greens","gourds","kovaikai","kothavarai","Lfinger","malli","mango","mushroom","nuckol","onion","peas","potroot","paneer","potato","pumkin","pudina","radish","tomato","tondaikai","vazpooThandu","yams","gram","channa","dhal","maida","oats","ragi","rice","wheat","coconut","oil","garlic","ginger","pepper","tamarind","aginomoto","spices","coco","horlicks","boost","nuts","runningnose","sneeze","cough","wheezeBlocks","headache","itching","swelling","redrashes","FHistory","Sex","Diagnosis")

> table(allergydataset2$Diagnosis)

**NORMAL OT RH RH\_O RH\_UT UT UT\_O**

**49 95 152 62 61 251 54**

CONSTANT SAMPLES

col\_names2 <- names(allergydataset2)

> allergydataset2[,col\_names] <- lapply(allergydataset2[,col\_names] , factor)

> ideal2<-SmoteClassif(Diagnosis ~ ., allergydataset2, C.perc ="balance",dist="HEOM")

> table(ideal2$Diagnosis)

**NORMAL OT RH RH\_O RH\_UT UT UT\_O**

**102 103 103 102 102 103 103**

BALANCING ONLY 4 SINGLETON CLASSES

singleton2<-SmoteClassif(Diagnosis ~ ., allergydataset2, C.perc =list(NORMAL=1.4,OT=1.2,RH=1.2,RH\_O=1,RH\_UT=1,UT=0.73,UT\_O=1),dist="HEOM")

table(singleton2$Diagnosis)

**NORMAL OT RH RH\_O RH\_UT UT UT\_O**

**68 113 182 62 61 183 54**

BALANCING ALL 7 CLASSES

full2<-SmoteClassif(Diagnosis ~ ., allergydataset2, C.perc =list(NORMAL=1.4,OT=1.2,RH=0.9,RH\_O=1.2,RH\_UT=1.5,UT=0.72,UT\_O=1.2),dist="HEOM")

> table(full2$Diagnosis)

**NORMAL OT RH RH\_O RH\_UT UT UT\_O**

**68 113 136 74 91 180 64**

write.csv(full2,"E:\\Ph.D\\3-JOURNALS\\A2-ALLERGY DSS\\ADSS Codes\\Allergy\_Data2\_full.csv")

> write.csv(singleton2,"E:\\Ph.D\\3-JOURNALS\\A2-ALLERGY DSS\\ADSS Codes\\Allergy\_Data2\_single.csv")

> write.csv(ideal2,"E:\\Ph.D\\3-JOURNALS\\A2-ALLERGY DSS\\ADSS Codes\\Allergy\_Data2\_ideal.csv")

>

> dim(ideal2)

**[1] 718 91**

> dim(singleton2)

**[1] 723 91**

> dim(full2)

**[1] 726 91**

**DATA SAMPLING PLOTS CODE**

dim(allergydataset)

[1] 904 92

> table(ubtrain$Diagnosis)

NORMAL OT RH RH\_O RH\_UT UT UT\_O

47 101 152 64 56 253 52

> plot(ubtrain$Diagnosis)

> plot(ubtrain$Diagnosis, font=2)

> table(sconstant$Diagnosis)

NORMAL OT RH RH\_O RH\_UT UT UT\_O

104 104 104 104 104 104 104

> plot(sconstant$Diagnosis, font=2)

> table(ssingleton$Diagnosis)

NORMAL OT RH RH\_O RH\_UT UT UT\_O

65 141 167 64 56 184 52

> plot(ssingleton$Diagnosis, font=2)

> table(sfull$Diagnosis)

NORMAL OT RH RH\_O RH\_UT UT UT\_O

42 90 144 83 95 189 83

> plot(sfull$Diagnosis, font=2)