Data Mining Lab (CSEN 2151)

On Decision Tree Classification

1. **Execute each line one by one, and list down the values of each of the variables. What are the role of these variables?**

```
d <- read.csv("ClassificationSimpleLab.csv")
d
f <- table(d$buys_computer)
f

t<-table(d$age, d$buys_computer)
t

prop1<-t[1,]/sum(t[1,])
prop1


prop2<-t[2,]/sum(t[2,])
prop2

prop3<-t[3,]/sum(t[3,])
prop3

H1 <- -(prop1[1]*log2(prop1[1]))-(prop1[2]*log2(prop1[2]))
H1
H11<-ifelse(is.na(H1),0,H1)
H11

H2<--(prop2[1]*log2(prop2[1]))-(prop2[2]*log2(prop2[2]))
H2
H22<-ifelse(is.na(H2),0,H2)
H22

H3 <- -(prop3[1]*log2(prop3[1]))-(prop3[2]*log2(prop3[2]))
H3
H33<-ifelse(is.na(H3),0,H3)
H33

x<-(table(d$age)[1]/length(d$age))*H11 +(table(d$age)[2]/length(d$age))*H22 +
  (table(d$age)[3]/length(d$age))*H33

xx<-ifelse(is.na(x),0,x)
xx <- unname(xx)
xx

r <- -(f[1]/nrow(d))*log2((f[1]/nrow(d)))-(f[2]/nrow(d))*log2((f[2]/nrow(d)))
r <- unname(r)

e <-r –xx
e
```

2. **Compute the values of xx, r and e for each of the other effective input variables in the given file.**
3. **Using Gini Index, Compute the values of xx, r and e for each of the effective input variables in the given file.**
4. **Write a generalized function Compute(x,y) to display values of xx,r and e for any x with respect to y ,here x is input variable column and y is the class variable column using entropy.**