12.11 using Logistic Regression, Linear Discriminant Analysis, kNN, and Naïve Bayes Classifications as well as two ensembles of these learners (Committees and Weighted Average). Evaluate for the following new entrant:

(age=34, education=finished high school, income=20, depression=normal, regular drinking=yes)

<u>Logistic Regression:</u>

<u>Code:</u>

```
data(Depression)

#Logistic regression model
model<-glm(formula=chronic~AGE+EDUCAT+INCOME+Depressed+regular_drinker, data = Depression, family = "binomial")
model
summary(model)
model_opt<-step(model,direction = 'backward',scope=formula(model))
summary(model_opt)

#testing given values
given_data1<-c(34,3,20,0,1)
given_data2<-c('AGE','EDUCAT','INCOME','Depressed','regular_drinker')

data1<-data.frame(row.names=given_data2,given_data1)
data1<-as.data.frame(t(data1))

Testing_values<-predict(model_opt,data1,type="response")
Testing_values
```

<u>Output:</u>

```
Call:  glm(formula = chronic ~ AGE + EDUCAT + INCOME + Depressed + regular_drinker,
    family = "binomial", data = Depression)

Coefficients:
    (Intercept)              AGE           EDUCAT           INCOME        Depressed  regular_drinker
      -1.844003         0.027554         0.006797        -0.001006         0.737405         0.437957

Degrees of Freedom: 293 Total (i.e. Null);  288 Residual
Null Deviance:      407.5
Residual Deviance: 384.2        AIC: 396.2
```

```
> summary(model)

Call:
glm(formula = chronic ~ AGE + EDUCAT + INCOME + Depressed + regular_drinker,
    family = "binomial", data = Depression)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8022  -1.0893   0.6484   1.0944   1.5862

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.844003   0.657723  -2.804 0.005053 **
AGE              0.027554   0.007260   3.795 0.000148 ***
EDUCAT           0.006797   0.103482   0.066 0.947629
INCOME          -0.001006   0.008921  -0.113 0.910258
Depressed        0.737405   0.339473   2.172 0.029840 *
regular_drinker  0.437957   0.309365   1.416 0.156874
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 407.52  on 293  degrees of freedom
Residual deviance: 384.19  on 288  degrees of freedom
AIC: 396.19

Number of Fisher Scoring iterations: 4
```

```
> model_opt<-step(model,direction = 'backward',scope=formula(model))
Start:  AIC=396.19
chronic ~ AGE + EDUCAT + INCOME + Depressed + regular_drinker

                  Df Deviance    AIC
- EDUCAT           1   384.20 394.20
- INCOME           1   384.21 394.21
<none>                 384.19 396.19
- regular_drinker  1   386.22 396.22
- Depressed        1   389.07 399.07
- AGE              1   399.45 409.45

Step:  AIC=394.2
chronic ~ AGE + INCOME + Depressed + regular_drinker

                  Df Deviance    AIC
- INCOME           1   384.21 392.21
<none>                 384.20 394.20
- regular_drinker  1   386.22 394.22
- Depressed        1   389.08 397.08
- AGE              1   399.71 407.71

Step:  AIC=392.21
chronic ~ AGE + Depressed + regular_drinker

                  Df Deviance    AIC
<none>                 384.21 392.21
- regular_drinker  1   386.26 392.26
- Depressed        1   389.36 395.36
- AGE              1   400.50 406.50
```

```
> summary(model_opt)

Call:
glm(formula = chronic ~ AGE + Depressed + regular_drinker, family = "binomial",
    data = Depression)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8030  -1.0857   0.6464   1.0959   1.5799

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.846130   0.483788  -3.816 0.000136 ***
AGE              0.027612   0.007049   3.917 8.95e-05 ***
Depressed        0.742119   0.332332   2.233 0.025545 *
regular_drinker  0.439350   0.308350   1.425 0.154203
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 407.52  on 293  degrees of freedom
Residual deviance: 384.21  on 290  degrees of freedom
AIC: 392.21

Number of Fisher Scoring iterations: 4


> Testing_values
given_data1
  0.3850981
```

The predicted value for chronic illness is 0.3850981 that is close to zero => No chronic illness for the last year for the given values.

KNN:

Code:

```
library(class)

#KNN model
Depress_knn<-Depression[,c('AGE','EDUCAT','INCOME','Depressed','regular_drinker','chronic')]

#variable name that stores the variable we need to classify
abso<-Knn_model$chronic


#Taking k=5
predict_1<-knn(train=Knn_model[,-6],test=data1,cl=abso,k=5)
predict_1

#taking k=9
predict_2<-knn(train=Knn_model[,-6],test=data1,cl=abso,k=9)
predict_2
```

Output:

```
> predict_1
[1] 1
Levels: 0 1
```

```
> predict_2
[1] 1
Levels: 0 1
```

=>Checking with k values 5 and 9 we get the output as 1=> Chronic illness is there last year

Naïve Bayes:

Code:

```
library(e1071)

#Naive Beyes
Depress_beyes<-Depression[,c('AGE','EDUCAT','INCOME','Depressed','regular_drinker','chronic')]
model_beyes<-naiveBayes(as.factor(chronic)~.,data=Depress_beyes)

predict_3<-predict(model_beyes,data1)
predict_3
```

Output:

```
> predict_3
[1] 0
Levels: 0 1
```

=>We get the value 0=> No Chronic illness last year.

Linear Discriminant Analysis:

Code:

```
library(MASS)

#Linear Discriminant Analysis
lda_mod<- lda(formula = chronic ~ AGE+EDUCAT+INCOME+Depressed+regular_drinker, data = Depression)
lda_mod

predict_4 <- predict(lda_mod, data1)
predict_4

predict_4$class
```

Output:

```
> lda_mod
Call:
lda(chronic ~ AGE + EDUCAT + INCOME + Depressed + regular_drinker,
    data = Depression)

Prior probabilities of groups:
        0         1
0.4931973 0.5068027

Group means:
       AGE   EDUCAT   INCOME Depressed regular_drinker
0 40.15862 3.572414 21.79310 0.1310345        1.158621
1 48.55705 3.389262 19.38926 0.2080537        1.248322

Coefficients of linear discriminants:
                         LD1
AGE              0.048391793
EDUCAT           0.010179285
INCOME          -0.001791659
Depressed        1.271536596
regular_drinker  0.751784398
```

```
> predict_4
$class
[1] 0
Levels: 0 1

$posterior
                    0         1
given_data1 0.6164345 0.3835655

$x
                  LD1
given_data1 -0.877524

> predict_4$class
[1] 0
Levels: 0 1
```

= We get the value as 0 => No Chronic illness for last year.

Ensemble:

Code:

```
p<-as.numeric(Testing_values)
q<-as.numeric(predict_2)
r<-as.numeric(predict_3)
t<-as.numeric(predict_4$class)

library(Cubist)
library(mlbench)


#commitees
set.seed(1)

sampling1 = sample(1:nrow(Depress_beyes),as.integer(0.8*nrow(Depress_beyes)))

training= Depress_beyes[sampling1,]
testing  = Depress_beyes[sampling1,]

training_o = Depress_beyes$chronic[sampling1]
testing_o = Depress_beyes$chronic[-sampling1]

set.seed(1)

ensemble_comi <- cubist(x=training,y=training_o,committees=4)
summary(ensemble_comi)

#weighted_average
weighted_a <- c(p,q,r,t)
weighted_a1<-mean(weighted_a)
weighted_a1
```

Output:

```
> summary(ensemble_comi)

Call:
cubist.default(x = training, y = training_o, committees = 4)


Cubist [Release 2.07 GPL Edition]  Thu Oct 18 18:01:53 2018
---------------------------------

    Target attribute `outcome'

Read 235 cases (7 attributes) from undefined.data

Model 1:

  Rule 1/1: [235 cases, mean 0.5, range 0 to 1, est err 0.0]

        outcome = 0 + 1 chronic

Model 2:

  Rule 2/1: [235 cases, mean 0.5, range 0 to 1, est err 0.0]

        outcome = 0 + 1 chronic

Model 3:

  Rule 3/1: [235 cases, mean 0.5, range 0 to 1, est err 0.0]

        outcome = 0 + 1 chronic



Model 4:

  Rule 4/1: [235 cases, mean 0.5, range 0 to 1, est err 0.0]

        outcome = 0 + 1 chronic


Evaluation on training data (235 cases):

    Average  |error|              0.0
    Relative |error|             0.00
    Correlation coefficient      1.00


        Attribute usage:
          Conds  Model

                 100%    chronic


Time: 0.0 secs
```

```
> weighted_a1
[1] 1.096275
```

=>Ensemble (committee) we get the value as 100 percent chronic illness in last year.

=>Ensemble (weighted average) we get the value as 1.096275 that is 1 => chronic illness is there last year.