Experiment 5

AIM: Write a program to perform weather forecasting using R

Code:

```
df<-read.csv('weather.csv')
apply(apply(df,2,is.na),2,sum)
df<-na.omit(df)
binca<-df$RainTomorrow
df$RainTomorrow[df$RainTomorrow =='Yes']<-1
df$RainTomorrow[df$RainTomorrow =='No']<-0
set.seed(123)
split = sample.split(df$RainTomorrow, SplitRatio = 0.8)
training_set = subset(df, split == TRUE)
test_set = subset(df, split == FALSE)
training_set<-training_set[-c(1,2,8,10,11,23)]
mytrainset2<-training_set[-18]
training_set<-lapply(training_set,as.numeric)
model1<-
glm(training_set$RainTomorrow~.,data=mytrainset2,family=binomial())
summary(model1)
test_set<-test_set[-c(1,2,8,10,11,23)]
mytestset2<-test_set[-18]
test_set<-lapply(test_set,as.numeric)
predi1<-predict(model1,mytestset2,type="response")
y_pred = ifelse(predi1 > 0.5, 1, 0)
tab1 <- table(Predicted = predi1, Actual = test_set$RainTomorrow)
missing_classerr <- mean(y_pred != test_set$RainTomorrow)
print(paste('Accuracy =', 1 - missing_classerr))
exp(coef(model1))
```

```
anova(model1,test="Chisq")

model2<-
glm(training_set$RainTomorrow~MinTemp+MaxTemp+Sunshine+WindGustS
peed+Humidity9am+Humidity3pm+Pressure3pm,data=mytrainset2,family=bin
omial())

summary(model2)

1-pchisq(deviance(model2)-deviance(model1),df.residual(model2)-
df.residual(model1))

predi2<-predict(model2,mytestset2,type="response")

binca<-ifelse(binca=="Yes",c(1),c(0))

summ=0

for(i in 1:length(predi1)){summ=summ+(binca[i]-predi1[i])^2}

summ


summ=0

for(i in 1:length(predi2)){summ=summ+(binca[i]-predi2[i])^2}

summ

mean(binca)

sum=0

for(i in 1:length(binca)){sum=sum+(binca[i]-mean(binca))^2}

sum
```

Output:

```
> df<-read.csv('weather.csv')
> apply(apply(df,2,is.na),2,sum)
        Date      Location       MinTemp       MaxTemp       Rainfall   Evaporation      Sunshine
           0             0             0             0             0             0             3
  WindGustDir WindGustSpeed     WindDir9am     WindDir3pm  WindSpeed9am  WindSpeed3pm    Humidity9am
           3             2            31             1             7             0             0
  Humidity3pm    Pressure9am    Pressure3pm      Cloud9am      Cloud3pm       Temp9am        Temp3pm
           0             0             0             0             0             0             0
    RainToday       RISK_MM  RainTomorrow
           0             0             0
> df<-na.omit(df)
> binca<-df$RainTomorrow
> df$RainTomorrow[df$RainTomorrow =='Yes']<-1
> df$RainTomorrow[df$RainTomorrow =='No']<-0
> set.seed(123)
> split = sample.split(df$RainTomorrow, SplitRatio = 0.8)
Error in sample.split(df$RainTomorrow, SplitRatio = 0.8) :
  could not find function "sample.split"
> training_set = subset(df, split == TRUE)
> test_set = subset(df, split == FALSE)
> training_set<-training_set[-c(1,2,8,10,11,23)]
> mytrainset2<-training_set[-18]
> training_set<-lapply(training_set,as.numeric)
Warning message:
In lapply(training_set, as.numeric) : NAs introduced by coercion
> model1<-glm(training_set$RainTomorrow~.,data=mytrainset2,family=binomial())
> summary(model1)

Call:
glm(formula = training_set$RainTomorrow ~ ., family = binomial(),
    data = mytrainset2)

Deviance Residuals:
    Min       1Q     Median       3Q       Max
-2.14927  -0.38430  -0.18007  -0.06652   2.43614

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   123.625886  54.710756   2.260 0.023844 *
MinTemp        -0.207058   0.158639  -1.305 0.191820
MaxTemp        -0.028772   0.266982  -0.108 0.914181
Rainfall       -0.037974   0.082738  -0.459 0.646255
Evaporation     0.149512   0.139944   1.068 0.285351
Sunshine       -0.250267   0.154959  -1.615 0.106299
WindGustSpeed   0.019025   0.031520   0.604 0.546122
WindSpeed9am    0.075650   0.049791   1.519 0.128669
WindSpeed3pm    0.007915   0.040858   0.194 0.846398
Humidity9am     0.054637   0.036961   1.478 0.139344
Humidity3pm     0.065631   0.038055   1.725 0.084594 .
Pressure9am     0.686808   0.235642   2.915 0.003561 **
Pressure3pm    -0.824781   0.242893  -3.396 0.000685 ***
Cloud9am        0.044568   0.134685   0.331 0.740717
Cloud3pm        0.154399   0.143295   1.077 0.281261
Temp9am         0.126304   0.235915   0.535 0.592388
Temp3pm         0.229848   0.281252   0.817 0.413796
RainTodayYes    0.737751   0.890194   0.829 0.407244
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
```

```
RainTodayYes     0.737751    0.890194    0.829 0.407244
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 249.54  on 261  degrees of freedom
Residual deviance: 125.18  on 244  degrees of freedom
AIC: 161.18

Number of Fisher Scoring iterations: 7

> test_set<-test_set[-c(1,2,8,10,11,23)]
> mytestset2<-test_set[-18]
> test_set<-lapply(test_set,as.numeric)
Warning message:
In lapply(test_set, as.numeric) : NAs introduced by coercion
> predi1<-predict(model1,mytestset2,type="response")
> y_pred = ifelse(predi1 > 0.5, 1, 0)
> tab1 <- table(Predicted = predi1, Actual = test_set$RainTomorrow)
> missing_classerr <- mean(y_pred != test_set$RainTomorrow)
> print(paste('Accuracy =', 1 - missing_classerr))
[1] "Accuracy = 0.863636363636364"
> exp(coef(model1))
  (Intercept)       MinTemp        MaxTemp       Rainfall    Evaporation       Sunshine WindGustSpeed
 4.898238e+53  8.129726e-01   9.716381e-01   9.627377e-01   1.161268e+00   7.785930e-01  1.019207e+00
  WindSpeed9am   WindSpeed3pm     Humidity9am     Humidity3pm     Pressure9am     Pressure3pm       Cloud9am
 1.078585e+00  1.007946e+00   1.056157e+00   1.067833e+00   1.987361e+00   4.383310e-01  1.045576e+00
      Cloud3pm        Temp9am         Temp3pm     RainTodayYes
 1.166957e+00  1.134627e+00   1.258409e+00   2.091226e+00
> anova(model1,test="Chisq")
Analysis of Deviance Table

Model: binomial, link: logit

Response: training_set$RainTomorrow

Terms added sequentially (first to last)


              Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                           261     249.54
MinTemp        1   12.179       260     237.36 0.0004834 ***
MaxTemp        1    8.516       259     228.84 0.0035200 **
Rainfall       1    0.382       258     228.46 0.5366904
Evaporation    1    0.384       257     228.08 0.5352358
Sunshine       1   32.168       256     195.91 1.414e-08 ***
WindGustSpeed  1   23.617       255     172.29 1.175e-06 ***
WindSpeed9am   1    0.015       254     172.28 0.9039217
WindSpeed3pm   1    0.020       253     172.26 0.8868929
Humidity9am    1   14.981       252     157.28 0.0001086 ***
Humidity3pm    1    4.125       251     153.15 0.0422614 *
Pressure9am    1    5.964       250     147.19 0.0145968 *
Pressure3pm    1   18.683       249     128.50 1.544e-05 ***
Cloud9am       1    0.035       248     128.47 0.8523197
Cloud3pm       1    1.240       247     127.23 0.2655565
Temp9am        1    0.561       246     126.67 0.4536685
Temp3pm        1    0.815       245     125.86 0.3667633
RainToday      1    0.676       244     125.18 0.4109843
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model2<-glm(training_set$RainTomorrow~MinTemp+MaxTemp+Sunshine+WindGustSpeed+Humidity9am+Humidity3pm+Pressure3
```

```
> model2<-glm(training_set$RainTomorrow~MinTemp+MaxTemp+Sunshine+WindGustSpeed+Humidity9am+Humidity3pm+Pressure
())
> summary(model2)

Call:
glm(formula = training_set$RainTomorrow ~ MinTemp + MaxTemp +
    Sunshine + WindGustSpeed + Humidity9am + Humidity3pm + Pressure3pm,
    family = binomial(), data = mytrainset2)

Deviance Residuals:
    Min       1Q    Median       3Q      Max
-2.08109  -0.44194  -0.22296  -0.09943   2.44321

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)   148.23544   50.24751   2.950 0.003177 **
MinTemp        -0.18863    0.07654  -2.464 0.013721 *
MaxTemp         0.28686    0.08491   3.379 0.000729 ***
Sunshine       -0.41413    0.11352  -3.648 0.000264 ***
WindGustSpeed   0.04813    0.02109   2.282 0.022474 *
Humidity9am     0.04856    0.02212   2.195 0.028157 *
Humidity3pm     0.02976    0.02473   1.203 0.228817
Pressure3pm    -0.15605    0.04862  -3.210 0.001328 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 249.54  on 261  degrees of freedom
Residual deviance: 142.88  on 254  degrees of freedom
AIC: 158.88

Number of Fisher Scoring iterations: 6

> 1-pchisq(deviance(model2)-deviance(model1),df.residual(model2)-df.residual(model1))
[1] 0.06023695
> predi2<-predict(model2,mytestset2,type="response")
> binca<-ifelse(binca=="Yes",c(1),c(0))
> summ=0
> for(i in 1:length(predi1)){summ=summ+(binca[i]-predi1[i])^2}
> summ
       4
20.91554
> summ=0
> for(i in 1:length(predi2)){summ=summ+(binca[i]-predi2[i])^2}
> summ
       4
20.19355
> mean(binca)
[1] 0.1829268
> sum=0
> for(i in 1:length(binca)){sum=sum+(binca[i]-mean(binca))^2}
> sum
[1] 49.02439
>
```