Date: 10/April/2024 EXPERIMENT – 10	K- NEAREST NEIGBOR CLASSIFIER

AIM: To perform K nearest neighbor classifier

SOFTWARE REQUIRED: RStudio

R CODE:

```
rm(list=ls())
data <- read.csv ("CreditWorthiness (1).csv", stringsAsFactors = TRUE) #This line reads the data from the CSV file named "CreditWorthiness.csv" into a data frame called data, with strings converted to factors.
str(data)
summary(data)
plot(data)
data$Cdur <-as.integer(data$Cdur)
data$Cpur <-as.integer(data$Cpur)
data$Camt <- as.integer(data$Camt)
data$age <- as.integer(data$age)
data[, -5] <- scale(data[, -5])
set.seed(123)
train_indices <- sample (nrow (data), 900)
data_TRAIN <- data[train_indices, ]
data_TEST <- data[-train_indices, ]
library (class)
library (caret)
knnpredict <- knn(train = data_TRAIN[, -5], test= data_TEST[,- 5], cl = data_TRAIN$creditScore, k = 5)
confusionMatrix(table(knnpredict, data_TEST$creditScore), positive = 'good')
```

OUTPUT:

```
> str(data)
'data.frame':   1000 obs. of  5 variables:
 $ Cdur       : int  9 15 36 48 24 27 12 12 36 36 ...
 $ Cpur       : Factor w/ 10 levels "Business","domestic needs",..: 1 4 1 1 4 8 4 10 3 4 ...
 $ Camt       : int  13790 15250 19410 144090 31690 51780 21590 9950 18070 23820 ...
 $ age        : int  27 50 61 25 26 48 29 22 37 25 ...
 $ creditScore: Factor w/ 2 levels "bad","good": 2 2 1 1 2 2 2 2 1 2 ...
> #This line displays the structure of the data dataframe, showing the data types and the first
e.
> summary(data)
```

e.

```
> summary(data)
      Cdur                      Cpur          Camt             age        creditScore
 Min.   : 4.0   electronics        :280   Min.   :  2380   Min.   :19.00   bad :300
 1st Qu.:12.0   second hand vehicle:234   1st Qu.: 13535   1st Qu.:27.00   good:700
 Median :18.0   furniture          :181   Median : 23075   Median :33.00
 Mean   :20.9   new vehicle        :103   Mean   : 32593   Mean   :35.55
 3rd Qu.:24.0   Business           : 97   3rd Qu.: 39603   3rd Qu.:42.00
 Max.   :72.0   education          : 50   Max.   :184120   Max.   :75.00
                (Other)            : 55
```

```
> confusionMatrix(table(knnpredict, data_test
Confusion Matrix and Statistics


knnpredict bad good
      bad    9   12
      good  21   58

               Accuracy : 0.67
                 95% CI : (0.5688, 0.7608)
    No Information Rate : 0.7
    P-Value [Acc > NIR] : 0.7793

                  Kappa : 0.1406

 Mcnemar's Test P-Value : 0.1637

            Sensitivity : 0.8286
            Specificity : 0.3000
         Pos Pred Value : 0.7342
         Neg Pred Value : 0.4286
             Prevalence : 0.7000
         Detection Rate : 0.5800
   Detection Prevalence : 0.7900
```

```
Mcnemar's Test P-Value : 0.1637

            Sensitivity : 0.8286
            Specificity : 0.3000
         Pos Pred Value : 0.7342
         Neg Pred Value : 0.4286
             Prevalence : 0.7000
         Detection Rate : 0.5800
   Detection Prevalence : 0.7900
      Balanced Accuracy : 0.5643

       'Positive' Class : good

>
```