Classification-on-Crop\_recommendation.R

user

2023-05-12

library(caTools)  
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.2.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.3

library(caret)

## Warning: package 'caret' was built under R version 4.2.3

## Loading required package: lattice

library(class)

## Warning: package 'class' was built under R version 4.2.3

library(corrplot)

## Warning: package 'corrplot' was built under R version 4.2.3

## corrplot 0.92 loaded

Crop\_recommendation <- read.csv("Crop\_recommendation.csv")  
  
  
  
names(Crop\_recommendation)

## [1] "N" "P" "K" "temperature" "humidity"   
## [6] "ph" "rainfall" "label"

head(Crop\_recommendation)

## N P K temperature humidity ph rainfall label  
## 1 90 42 43 20.87974 82.00274 6.502985 202.9355 rice  
## 2 85 58 41 21.77046 80.31964 7.038096 226.6555 rice  
## 3 60 55 44 23.00446 82.32076 7.840207 263.9642 rice  
## 4 74 35 40 26.49110 80.15836 6.980401 242.8640 rice  
## 5 78 42 42 20.13017 81.60487 7.628473 262.7173 rice  
## 6 69 37 42 23.05805 83.37012 7.073454 251.0550 rice

tail(Crop\_recommendation)

## N P K temperature humidity ph rainfall label  
## 1692 100 76 45 25.56703 75.94068 5.590236 102.78677 banana  
## 1693 117 86 48 28.69562 82.54196 6.225225 116.16168 banana  
## 1694 114 94 53 26.33545 76.85320 6.190757 118.68583 banana  
## 1695 110 78 50 25.93730 78.89864 5.915569 98.21748 banana  
## 1696 94 70 48 25.13687 84.88394 6.195152 91.46442 banana  
## 1697 80 71 47 27.50528 80.79784 6.156373 105.07770 banana

summary(Crop\_recommendation)

## N P K temperature   
## Min. : 0.00 Min. : 5.00 Min. : 5.00 Min. : 8.826   
## 1st Qu.: 21.00 1st Qu.: 36.00 1st Qu.: 19.00 1st Qu.:21.631   
## Median : 37.00 Median : 54.00 Median : 27.00 Median :24.910   
## Mean : 52.65 Mean : 58.13 Mean : 52.03 Mean :24.712   
## 3rd Qu.: 90.00 3rd Qu.: 72.00 3rd Qu.: 52.00 3rd Qu.:27.835   
## Max. :140.00 Max. :145.00 Max. :205.00 Max. :41.949   
## humidity ph rainfall label   
## Min. :14.26 Min. :3.505 Min. : 5.315 Length:1697   
## 1st Qu.:51.28 1st Qu.:5.861 1st Qu.: 66.839 Class :character   
## Median :77.91 Median :6.354 Median : 93.123 Mode :character   
## Mean :65.85 Mean :6.439 Mean : 99.365   
## 3rd Qu.:83.86 3rd Qu.:6.933 3rd Qu.:115.356   
## Max. :94.96 Max. :9.935 Max. :298.560

str(Crop\_recommendation)

## 'data.frame': 1697 obs. of 8 variables:  
## $ N : int 90 85 60 74 78 69 69 94 89 68 ...  
## $ P : int 42 58 55 35 42 37 55 53 54 58 ...  
## $ K : int 43 41 44 40 42 42 38 40 38 38 ...  
## $ temperature: num 20.9 21.8 23 26.5 20.1 ...  
## $ humidity : num 82 80.3 82.3 80.2 81.6 ...  
## $ ph : num 6.5 7.04 7.84 6.98 7.63 ...  
## $ rainfall : num 203 227 264 243 263 ...  
## $ label : chr "rice" "rice" "rice" "rice" ...

dim(Crop\_recommendation)

## [1] 1697 8

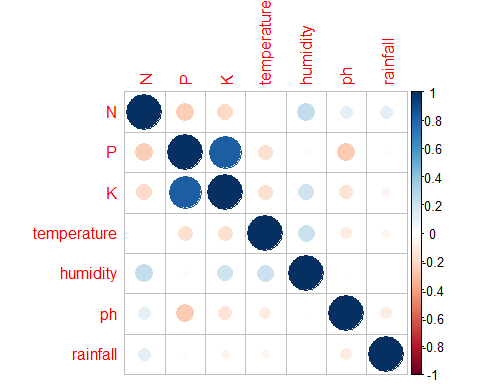
standard.features <- scale(Crop\_recommendation[,1:7])  
#Join the standardized data with the target column   
data <- cbind(standard.features,Crop\_recommendation[8])   
head(data)

## N P K temperature humidity ph rainfall  
## 1 0.9692765 -0.472015676 -0.1582517 -0.7803659 0.6611602 0.07355549 2.053489  
## 2 0.8395310 -0.003691126 -0.1932971 -0.5989728 0.5922728 0.68833514 2.523785  
## 3 0.1908032 -0.091501979 -0.1407289 -0.3476714 0.6741763 1.60986605 3.263503  
## 4 0.5540908 -0.676907667 -0.2108198 0.3623761 0.5856717 0.62204985 2.845150  
## 5 0.6578872 -0.472015676 -0.1757744 -0.9330143 0.6448757 1.36660831 3.238780  
## 6 0.4243452 -0.618367098 -0.1757744 -0.3367580 0.7171252 0.72895634 3.007552  
## label  
## 1 rice  
## 2 rice  
## 3 rice  
## 4 rice  
## 5 rice  
## 6 rice

anyNA(data)

## [1] FALSE

corrplot(cor(data[,-8]))



set.seed(101)  
  
sample <- sample.split(data$label,SplitRatio = 0.70)  
train <- subset(data,sample==TRUE)  
dim(train)

## [1] 1187 8

test <- subset(data,sample==FALSE)  
dim(test)

## [1] 510 8

predicted.type <- knn(train[1:7], test[1:7], train$label,k=1)  
#Error in prediction  
error <- mean(predicted.type!=test$label)  
#Confusion Matrix  
confusionMatrix(predicted.type,as.factor(test$label))

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction apple banana beans coffee cotton cowpeas grapes groundnuts maize  
## apple 30 0 0 0 0 0 0 0 0  
## banana 0 39 0 0 0 0 0 0 0  
## beans 0 0 37 0 0 0 0 0 0  
## coffee 0 0 0 33 0 0 0 0 0  
## cotton 0 0 0 0 29 0 0 0 2  
## cowpeas 0 0 0 0 0 37 0 0 0  
## grapes 0 0 0 0 0 0 30 0 0  
## groundnuts 0 0 0 0 0 0 0 30 0  
## maize 0 0 0 0 1 0 0 0 34  
## mango 0 0 0 0 0 0 0 0 0  
## orange 0 0 0 0 0 0 0 0 0  
## peas 0 0 0 0 0 0 0 0 0  
## rice 0 0 0 0 0 0 0 0 0  
## Soyabeans 0 0 0 0 0 0 0 0 0  
## watermelon 0 0 0 0 0 0 0 0 0  
## Reference  
## Prediction mango orange peas rice Soyabeans watermelon  
## apple 0 0 0 0 0 0  
## banana 0 0 0 0 0 0  
## beans 0 0 1 0 0 0  
## coffee 0 0 0 0 0 0  
## cotton 0 0 0 0 0 0  
## cowpeas 0 0 0 0 0 0  
## grapes 0 0 0 0 0 0  
## groundnuts 0 0 1 0 0 0  
## maize 0 0 0 0 0 0  
## mango 30 0 0 0 0 0  
## orange 0 37 0 0 0 0  
## peas 0 0 28 0 0 0  
## rice 0 0 0 42 0 0  
## Soyabeans 0 0 0 0 39 0  
## watermelon 0 0 0 0 0 30  
##   
## Overall Statistics  
##   
## Accuracy : 0.9902   
## 95% CI : (0.9773, 0.9968)  
## No Information Rate : 0.0824   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.9895   
##   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: apple Class: banana Class: beans Class: coffee  
## Sensitivity 1.00000 1.00000 1.00000 1.00000  
## Specificity 1.00000 1.00000 0.99789 1.00000  
## Pos Pred Value 1.00000 1.00000 0.97368 1.00000  
## Neg Pred Value 1.00000 1.00000 1.00000 1.00000  
## Prevalence 0.05882 0.07647 0.07255 0.06471  
## Detection Rate 0.05882 0.07647 0.07255 0.06471  
## Detection Prevalence 0.05882 0.07647 0.07451 0.06471  
## Balanced Accuracy 1.00000 1.00000 0.99894 1.00000  
## Class: cotton Class: cowpeas Class: grapes  
## Sensitivity 0.96667 1.00000 1.00000  
## Specificity 0.99583 1.00000 1.00000  
## Pos Pred Value 0.93548 1.00000 1.00000  
## Neg Pred Value 0.99791 1.00000 1.00000  
## Prevalence 0.05882 0.07255 0.05882  
## Detection Rate 0.05686 0.07255 0.05882  
## Detection Prevalence 0.06078 0.07255 0.05882  
## Balanced Accuracy 0.98125 1.00000 1.00000  
## Class: groundnuts Class: maize Class: mango Class: orange  
## Sensitivity 1.00000 0.94444 1.00000 1.00000  
## Specificity 0.99792 0.99789 1.00000 1.00000  
## Pos Pred Value 0.96774 0.97143 1.00000 1.00000  
## Neg Pred Value 1.00000 0.99579 1.00000 1.00000  
## Prevalence 0.05882 0.07059 0.05882 0.07255  
## Detection Rate 0.05882 0.06667 0.05882 0.07255  
## Detection Prevalence 0.06078 0.06863 0.05882 0.07255  
## Balanced Accuracy 0.99896 0.97117 1.00000 1.00000  
## Class: peas Class: rice Class: Soyabeans Class: watermelon  
## Sensitivity 0.93333 1.00000 1.00000 1.00000  
## Specificity 1.00000 1.00000 1.00000 1.00000  
## Pos Pred Value 1.00000 1.00000 1.00000 1.00000  
## Neg Pred Value 0.99585 1.00000 1.00000 1.00000  
## Prevalence 0.05882 0.08235 0.07647 0.05882  
## Detection Rate 0.05490 0.08235 0.07647 0.05882  
## Detection Prevalence 0.05490 0.08235 0.07647 0.05882  
## Balanced Accuracy 0.96667 1.00000 1.00000 1.00000

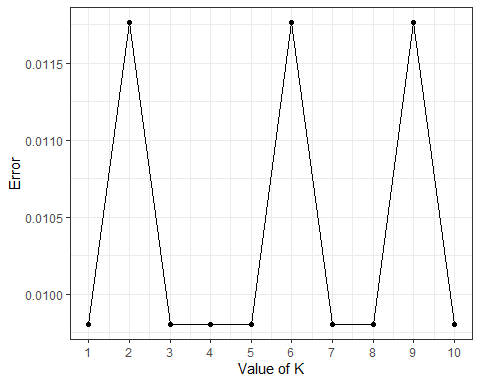
predicted.type <- NULL  
error.rate <- NULL  
for (i in 1:7) {  
 predicted.type <- knn(train[1:7],test[1:7],train$label,k=i)  
 error.rate[i] <- mean(predicted.type!=test$label)  
}  
knn.error <- as.data.frame(cbind(k=1:10,error.type =error.rate))

## Warning in cbind(k = 1:10, error.type = error.rate): number of rows of result  
## is not a multiple of vector length (arg 2)

knn.error

## k error.type  
## 1 1 0.009803922  
## 2 2 0.011764706  
## 3 3 0.009803922  
## 4 4 0.009803922  
## 5 5 0.009803922  
## 6 6 0.011764706  
## 7 7 0.009803922  
## 8 8 0.009803922  
## 9 9 0.011764706  
## 10 10 0.009803922

ggplot(knn.error,aes(k,error.type))+   
 geom\_point()+   
 geom\_line() +   
 scale\_x\_continuous(breaks=1:10)+   
 theme\_bw() +  
 xlab("Value of K") +  
 ylab('Error')



#step 14  
predicted.type <- knn(train[1:7],test[1:7],train$label,k=3)  
#Error in prediction  
error <- mean(predicted.type!=test$label)  
error

## [1] 0.01176471

#Confusion Matrix  
confusionMatrix(predicted.type,as.factor(test$label))

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction apple banana beans coffee cotton cowpeas grapes groundnuts maize  
## apple 30 0 0 0 0 0 0 0 0  
## banana 0 39 0 0 0 0 0 0 0  
## beans 0 0 37 0 0 0 0 0 0  
## coffee 0 0 0 33 0 0 0 0 0  
## cotton 0 0 0 0 29 0 0 0 2  
## cowpeas 0 0 0 0 0 37 0 0 0  
## grapes 0 0 0 0 0 0 30 0 0  
## groundnuts 0 0 0 0 0 0 0 30 0  
## maize 0 0 0 0 1 0 0 0 34  
## mango 0 0 0 0 0 0 0 0 0  
## orange 0 0 0 0 0 0 0 0 0  
## peas 0 0 0 0 0 0 0 0 0  
## rice 0 0 0 0 0 0 0 0 0  
## Soyabeans 0 0 0 0 0 0 0 0 0  
## watermelon 0 0 0 0 0 0 0 0 0  
## Reference  
## Prediction mango orange peas rice Soyabeans watermelon  
## apple 0 0 0 0 0 0  
## banana 0 0 0 0 0 0  
## beans 0 0 2 0 0 0  
## coffee 0 0 0 0 0 0  
## cotton 0 0 0 0 0 0  
## cowpeas 0 0 0 0 0 0  
## grapes 0 0 0 0 0 0  
## groundnuts 0 0 1 0 0 0  
## maize 0 0 0 0 0 0  
## mango 30 0 0 0 0 0  
## orange 0 37 0 0 0 0  
## peas 0 0 27 0 0 0  
## rice 0 0 0 42 0 0  
## Soyabeans 0 0 0 0 39 0  
## watermelon 0 0 0 0 0 30  
##   
## Overall Statistics  
##   
## Accuracy : 0.9882   
## 95% CI : (0.9746, 0.9957)  
## No Information Rate : 0.0824   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.9874   
##   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: apple Class: banana Class: beans Class: coffee  
## Sensitivity 1.00000 1.00000 1.00000 1.00000  
## Specificity 1.00000 1.00000 0.99577 1.00000  
## Pos Pred Value 1.00000 1.00000 0.94872 1.00000  
## Neg Pred Value 1.00000 1.00000 1.00000 1.00000  
## Prevalence 0.05882 0.07647 0.07255 0.06471  
## Detection Rate 0.05882 0.07647 0.07255 0.06471  
## Detection Prevalence 0.05882 0.07647 0.07647 0.06471  
## Balanced Accuracy 1.00000 1.00000 0.99789 1.00000  
## Class: cotton Class: cowpeas Class: grapes  
## Sensitivity 0.96667 1.00000 1.00000  
## Specificity 0.99583 1.00000 1.00000  
## Pos Pred Value 0.93548 1.00000 1.00000  
## Neg Pred Value 0.99791 1.00000 1.00000  
## Prevalence 0.05882 0.07255 0.05882  
## Detection Rate 0.05686 0.07255 0.05882  
## Detection Prevalence 0.06078 0.07255 0.05882  
## Balanced Accuracy 0.98125 1.00000 1.00000  
## Class: groundnuts Class: maize Class: mango Class: orange  
## Sensitivity 1.00000 0.94444 1.00000 1.00000  
## Specificity 0.99792 0.99789 1.00000 1.00000  
## Pos Pred Value 0.96774 0.97143 1.00000 1.00000  
## Neg Pred Value 1.00000 0.99579 1.00000 1.00000  
## Prevalence 0.05882 0.07059 0.05882 0.07255  
## Detection Rate 0.05882 0.06667 0.05882 0.07255  
## Detection Prevalence 0.06078 0.06863 0.05882 0.07255  
## Balanced Accuracy 0.99896 0.97117 1.00000 1.00000  
## Class: peas Class: rice Class: Soyabeans Class: watermelon  
## Sensitivity 0.90000 1.00000 1.00000 1.00000  
## Specificity 1.00000 1.00000 1.00000 1.00000  
## Pos Pred Value 1.00000 1.00000 1.00000 1.00000  
## Neg Pred Value 0.99379 1.00000 1.00000 1.00000  
## Prevalence 0.05882 0.08235 0.07647 0.05882  
## Detection Rate 0.05294 0.08235 0.07647 0.05882  
## Detection Prevalence 0.05294 0.08235 0.07647 0.05882  
## Balanced Accuracy 0.95000 1.00000 1.00000 1.00000