**Implement a KNN model to classify the animals in to categorie**

zoo<-read.csv(file.choose(),header = T)

View(zoo)

**#shows the structure of the data set**

str(zoo)

**#table of our target variable 'type**

table(zoo$type)

1 2 3 4 5 6 7

41 20 5 13 4 8 10

**#here we can see that all variable are in numerical format**

**#convert all those variable into text format for better understanding**

**# recode 'Type' as a factor**

**#following are the type of glasses**

**#1 -- Mammal**

**#2 --Bird**

**#3 --Reptiles**

**#4 --Fish**

**#5 --Frog**

**#6 --Insects**

**#7 --Marine**

zoo$type <- factor(zoo$type, levels = c(1,2,3,4,5,6,7),

labels = c("Mammal", "Bird","Reptiles","Fish","Frog","Insects","Marine"))

table(zoo$type)

table(zoo$type)

Mammal Bird Reptiles Fish Frog Insects Marine

41 20 5 13 4 8 10

View(zoo)

|  |
| --- |
| head(zoo)  animal.name hair feathers eggs milk airborne aquatic predator toothed backbone breathes  1 aardvark 1 0 0 1 0 0 1 1 1 1  2 antelope 1 0 0 1 0 0 0 1 1 1  3 bass 0 0 1 0 0 1 1 1 1 0  4 bear 1 0 0 1 0 0 1 1 1 1  5 boar 1 0 0 1 0 0 1 1 1 1  6 buffalo 1 0 0 1 0 0 0 1 1 1  venomous fins legs tail domestic catsize type  1 0 0 4 0 0 1 Mammal  2 0 0 4 1 0 1 Mammal  3 0 1 0 1 0 0 Fish  4 0 0 4 0 0 1 Mammal  5 0 0 4 1 0 1 Mammal  6 0 0 4 1 0 1 Mammal |
|  |
| |  | | --- | | > | |

str(zoo)

'data.frame': 101 obs. of 18 variables:

$ animal.name: chr "aardvark" "antelope" "bass" "bear" ...

$ hair : int 1 1 0 1 1 1 1 0 0 1 ...

$ feathers : int 0 0 0 0 0 0 0 0 0 0 ...

$ eggs : int 0 0 1 0 0 0 0 1 1 0 ...

$ milk : int 1 1 0 1 1 1 1 0 0 1 ...

$ airborne : int 0 0 0 0 0 0 0 0 0 0 ...

$ aquatic : int 0 0 1 0 0 0 0 1 1 0 ...

$ predator : int 1 0 1 1 1 0 0 0 1 0 ...

$ toothed : int 1 1 1 1 1 1 1 1 1 1 ...

$ backbone : int 1 1 1 1 1 1 1 1 1 1 ...

$ breathes : int 1 1 0 1 1 1 1 0 0 1 ...

$ venomous : int 0 0 0 0 0 0 0 0 0 0 ...

$ fins : int 0 0 1 0 0 0 0 1 1 0 ...

$ legs : int 4 4 0 4 4 4 4 0 0 4 ...

$ tail : int 0 1 1 0 1 1 1 1 1 0 ...

$ domestic : int 0 0 0 0 0 0 1 1 0 1 ...

$ catsize : int 1 1 0 1 1 1 1 0 0 0 ...

$ type : Factor w/ 7 levels "Mammal","Bird",..: 1 1 4 1 1 1 1 4 4 1 ...

**#normalize the data using customized normalized function**

normalize <- function(x) {

return ((x - min(x)) / (max(x) - min(x)))

}

zoo\_n<-as.data.frame(lapply(zoo[2:17],normalize))

View(zoo\_n)

head(zoo\_n)

hair feathers eggs milk airborne aquatic predator toothed backbone breathes venomous fins

1 1 0 0 1 0 0 1 1 1 1 0 0

2 1 0 0 1 0 0 0 1 1 1 0 0

3 0 0 1 0 0 1 1 1 1 0 0 1

4 1 0 0 1 0 0 1 1 1 1 0 0

5 1 0 0 1 0 0 1 1 1 1 0 0

6 1 0 0 1 0 0 0 1 1 1 0 0

legs tail domestic catsize

1 0.5 0 0 1

2 0.5 1 0 1

3 0.0 1 0 0

4 0.5 0 0 1

5 0.5 1 0 1

6 0.5 1 0 1

summary(zoo\_n)

hair feathers eggs milk airborne

Min. :0.0000 Min. :0.000 Min. :0.0000 Min. :0.0000 Min. :0.0000

1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000

Median :0.0000 Median :0.000 Median :1.0000 Median :0.0000 Median :0.0000

Mean :0.4257 Mean :0.198 Mean :0.5842 Mean :0.4059 Mean :0.2376

3rd Qu.:1.0000 3rd Qu.:0.000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000

Max. :1.0000 Max. :1.000 Max. :1.0000 Max. :1.0000 Max. :1.0000

aquatic predator toothed backbone breathes

Min. :0.0000 Min. :0.0000 Min. :0.000 Min. :0.0000 Min. :0.0000

1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:1.0000 1st Qu.:1.0000

Median :0.0000 Median :1.0000 Median :1.000 Median :1.0000 Median :1.0000

Mean :0.3564 Mean :0.5545 Mean :0.604 Mean :0.8218 Mean :0.7921

3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.000 3rd Qu.:1.0000 3rd Qu.:1.0000

Max. :1.0000 Max. :1.0000 Max. :1.000 Max. :1.0000 Max. :1.0000

venomous fins legs tail domestic

Min. :0.00000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000

1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.2500 1st Qu.:0.0000 1st Qu.:0.0000

Median :0.00000 Median :0.0000 Median :0.5000 Median :1.0000 Median :0.0000

Mean :0.07921 Mean :0.1683 Mean :0.3552 Mean :0.7426 Mean :0.1287

3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:0.5000 3rd Qu.:1.0000 3rd Qu.:0.0000

Max. :1.00000 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000

catsize

Min. :0.0000

1st Qu.:0.0000

Median :0.0000

Mean :0.4356

3rd Qu.:1.0000

Max. :1.0000

**#SPLITTING THE DATA SET**

**#here randomly select testing and training data set**

data<-sample(1:nrow(zoo\_n),size=nrow(zoo\_n)\*0.7,replace = FALSE)

**#Training and testing the dataset**

zoo\_train<-zoo\_n[data,]

zoo\_test<-zoo\_n[-data,]

**#creating training and testing labels**

zoo\_train\_labels<-zoo[data,18]

zoo\_test\_labels<-zoo[-data,18]

**#display first few rows of the traing and testing data set**

head(zoo\_train,10)

hair feathers eggs milk airborne aquatic predator toothed backbone breathes venomous fins

77 0 0 0 0 0 1 1 1 1 0 1 0

87 0 0 1 0 0 1 1 1 1 0 1 1

92 0 0 1 0 0 0 1 1 1 1 0 0

44 0 1 1 0 1 0 0 0 1 1 0 0

72 0 1 1 0 0 0 1 0 1 1 0 0

38 0 1 1 0 1 0 1 0 1 1 0 0

24 0 1 1 0 1 0 0 0 1 1 0 0

80 0 1 1 0 1 1 1 0 1 1 0 0

75 1 0 0 1 0 1 1 1 1 1 0 1

100 0 0 1 0 0 0 0 0 0 1 0 0

legs tail domestic catsize

77 0.00 1 0 0

87 0.00 1 0 1

92 0.50 1 0 0

44 0.25 1 0 0

72 0.25 1 0 1

38 0.25 1 0 0

24 0.25 1 0 1

80 0.25 1 0 0

75 0.00 0 0 1

100 0.00 0 0 0

head(zoo\_test,10)

hair feathers eggs milk airborne aquatic predator toothed backbone breathes venomous fins

5 1 0 0 1 0 0 1 1 1 1 0 0

7 1 0 0 1 0 0 0 1 1 1 0 0

8 0 0 1 0 0 1 0 1 1 0 0 1

9 0 0 1 0 0 1 1 1 1 0 0 1

11 1 0 0 1 0 0 1 1 1 1 0 0

17 0 1 1 0 1 0 1 0 1 1 0 0

18 1 0 0 1 0 0 0 1 1 1 0 0

19 0 0 1 0 0 1 1 1 1 0 0 1

20 0 0 0 1 0 1 1 1 1 1 0 1

22 0 1 1 0 1 1 0 0 1 1 0 0

legs tail domestic catsize

5 0.50 1 0 1

7 0.50 1 1 1

8 0.00 1 1 0

9 0.00 1 0 0

11 0.50 1 0 1

17 0.25 1 0 0

18 0.50 1 0 1

19 0.00 1 0 1

20 0.00 1 0 1

22 0.25 1 0 0

head(zoo\_train\_labels,10)

[1] Reptiles Fish Reptiles Bird Bird Bird Bird Bird Mammal Marine

Levels: Mammal Bird Reptiles Fish Frog Insects Marine

head(zoo\_test\_labels,10)

[1] Mammal Mammal Fish Fish Mammal Bird Mammal Fish Mammal Bird

Levels: Mammal Bird Reptiles Fish Frog Insects Marine

**#---- Training a model on the data ----**

# load the "class" library

install.packages("class")

library(class)

knn\_zoo\_pred1<-knn(train = zoo\_train,test = zoo\_test,cl = zoo\_train\_labels, k=1)

knn\_zoo\_pred1

[1] Mammal Mammal Fish Fish Mammal Bird Mammal Fish Mammal Bird

[11] Insects Mammal Insects Insects Mammal Mammal Reptiles Mammal Bird Fish

[21] Mammal Fish Bird Reptiles Mammal Frog Mammal Mammal Bird Mammal

[31] Mammal

Levels: Mammal Bird Reptiles Fish Frog Insects Marine

**#calculate the proportion of correct classification for k=1**

zoo\_accur<-100\*sum(zoo\_test\_labels==knn\_zoo\_pred1)/NROW(zoo\_test\_labels)

zoo\_accur

|  |
| --- |
| 96.77419 |
|  |
| |  | | --- | | > | |

**#table representation of predicted and actual data**

table(knn\_zoo\_pred,zoo\_test\_labels)

**##--------Evaluating model performance ----**

**# load the "gmodels" library**

library(gmodels)

CrossTable(x = zoo\_test\_labels, y = knn\_zoo\_pred)

Cell Contents

|-------------------------|

| N |

| Chi-square contribution |

| N / Row Total |

| N / Col Total |

| N / Table Total |

|-------------------------|

Total Observations in Table: 31

| knn\_zoo\_pred

zoo\_test\_labels | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Row Total |

----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

Mammal | 6 | 3 | 0 | 2 | 1 | 0 | 3 | 15 |

| 0.218 | 0.003 | 0.484 | 0.207 | 0.551 | 0.968 | 1.652 | |

| 0.400 | 0.200 | 0.000 | 0.133 | 0.067 | 0.000 | 0.200 | 0.484 |

| 0.400 | 0.500 | 0.000 | 0.667 | 1.000 | 0.000 | 1.000 | |

| 0.194 | 0.097 | 0.000 | 0.065 | 0.032 | 0.000 | 0.097 | |

----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

Bird | 4 | 0 | 1 | 0 | 0 | 0 | 0 | 5 |

| 1.033 | 0.968 | 4.361 | 0.484 | 0.161 | 0.323 | 0.484 | |

| 0.800 | 0.000 | 0.200 | 0.000 | 0.000 | 0.000 | 0.000 | 0.161 |

| 0.267 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | |

| 0.129 | 0.000 | 0.032 | 0.000 | 0.000 | 0.000 | 0.000 | |

----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

Reptiles | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

| 0.551 | 0.194 | 0.032 | 0.097 | 0.032 | 0.065 | 0.097 | |

| 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.032 |

| 0.067 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | |

| 0.032 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | |

----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

Fish | 2 | 1 | 0 | 1 | 0 | 1 | 0 | 5 |

| 0.073 | 0.001 | 0.161 | 0.551 | 0.161 | 1.423 | 0.484 | |

| 0.400 | 0.200 | 0.000 | 0.200 | 0.000 | 0.200 | 0.000 | 0.161 |

| 0.133 | 0.167 | 0.000 | 0.333 | 0.000 | 0.500 | 0.000 | |

| 0.065 | 0.032 | 0.000 | 0.032 | 0.000 | 0.032 | 0.000 | |

----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

Frog | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |

| 1.101 | 0.387 | 0.065 | 0.194 | 0.065 | 0.129 | 0.194 | |

| 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.065 |

| 0.133 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | |

| 0.065 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | |

----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

Insects | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 3 |

| 1.452 | 3.470 | 0.097 | 0.290 | 0.097 | 3.360 | 0.290 | |

| 0.000 | 0.667 | 0.000 | 0.000 | 0.000 | 0.333 | 0.000 | 0.097 |

| 0.000 | 0.333 | 0.000 | 0.000 | 0.000 | 0.500 | 0.000 | |

| 0.000 | 0.065 | 0.000 | 0.000 | 0.000 | 0.032 | 0.000 | |

----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

Column Total | 15 | 6 | 1 | 3 | 1 | 2 | 3 | 31 |

| 0.484 | 0.194 | 0.032 | 0.097 | 0.032 | 0.065 | 0.097 | |

----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

**#confusion matrix**

install.packages("caret")

library(caret)

confusionMatrix(table(knn\_zoo\_pred,zoo\_test\_labels))

|  |
| --- |
| Confusion Matrix and Statistics  zoo\_test\_labels  knn\_zoo\_pred1 Mammal Bird Reptiles Fish Frog Insects Marine  Mammal 15 0 0 0 0 0 0  Bird 0 5 0 0 0 0 0  Reptiles 0 0 1 0 1 0 0  Fish 0 0 0 5 0 0 0  Frog 0 0 0 0 1 0 0  Insects 0 0 0 0 0 3 0  Marine 0 0 0 0 0 0 0  Overall Statistics    Accuracy : 0.9677  95% CI : (0.833, 0.9992)  No Information Rate : 0.4839  P-Value [Acc > NIR] : 5.74e-09    Kappa : 0.9539    Mcnemar's Test P-Value : NA  Statistics by Class:  Class: Mammal Class: Bird Class: Reptiles Class: Fish Class: Frog  Sensitivity 1.0000 1.0000 1.00000 1.0000 0.50000  Specificity 1.0000 1.0000 0.96667 1.0000 1.00000  Pos Pred Value 1.0000 1.0000 0.50000 1.0000 1.00000  Neg Pred Value 1.0000 1.0000 1.00000 1.0000 0.96667  Prevalence 0.4839 0.1613 0.03226 0.1613 0.06452  Detection Rate 0.4839 0.1613 0.03226 0.1613 0.03226  Detection Prevalence 0.4839 0.1613 0.06452 0.1613 0.03226  Balanced Accuracy 1.0000 1.0000 0.98333 1.0000 0.75000  Class: Insects Class: Marine  Sensitivity 1.00000 NA  Specificity 1.00000 1  Pos Pred Value 1.00000 NA  Neg Pred Value 1.00000 NA  Prevalence 0.09677 0  Detection Rate 0.09677 0  Detection Prevalence 0.09677 0  Balanced Accuracy 1.00000 NA |
|  |
| |  | | --- | | > | |

**#using a simple loop we can find best k value**

i=1

k\_value=1

for(i in 1:25){

knn.model<-knn(train = zoo\_train, test = zoo\_test,cl = zoo\_train\_labels, k=i)

k\_value[i]<-100\*sum(zoo\_test\_labels == knn.model)/NROW(zoo\_test\_labels)

k=i

cat(k,"=",k\_value[i],'\n')

}

1 = 96.77419

2 = 100

3 = 100

4 = 96.77419

5 = 90.32258

6 = 93.54839

7 = 90.32258

8 = 90.32258

9 = 90.32258

10 = 93.54839

11 = 90.32258

12 = 87.09677

13 = 83.87097

14 = 83.87097

15 = 80.64516

16 = 80.64516

17 = 80.64516

18 = 80.64516

19 = 80.64516

20 = 80.64516

21 = 77.41935

22 = 80.64516

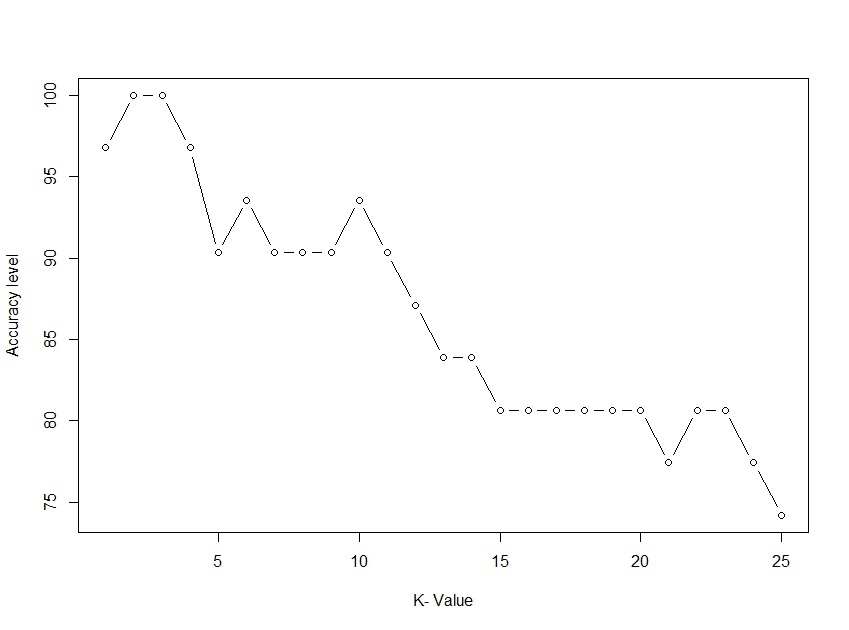
23 = 80.64516

24 = 77.41935

25 = 74.19355

**# plot the accuracy of k-value**

plot(k\_value, type="b", xlab="K- Value",ylab="Accuracy level")



**# create a dataframe contains, animal,test\_label, predicted data**

zoo\_animals<-zoo[-data,1]

test\_data<- -[data]

View(data.frame( zoo\_animals,zoo\_test\_labels,knn\_zoo\_pred1))

|  | **zoo\_animals** | **zoo\_test\_labels** | **knn\_zoo\_pred1** |
| --- | --- | --- | --- |
|  |  |  |  |
| **1** | boar | Mammal | Mammal |
| **2** | calf | Mammal | Mammal |
| **3** | carp | Fish | Fish |
| **4** | catfish | Fish | Fish |
| **5** | cheetah | Mammal | Mammal |
| **6** | crow | Bird | Bird |
| **7** | deer | Mammal | Mammal |
| **8** | dogfish | Fish | Fish |
| **9** | dolphin | Mammal | Mammal |
| **10** | duck | Bird | Bird |
| **11** | gnat | Insects | Insects |
| **12** | hamster | Mammal | Mammal |
| **13** | honeybee | Insects | Insects |
| **14** | ladybird | Insects | Insects |
| **15** | lion | Mammal | Mammal |
| **16** | mongoose | Mammal | Mammal |
| **17** | newt | Frog | Reptiles |
| **18** | opossum | Mammal | Mammal |
| **19** | parakeet | Bird | Bird |
| **20** | piranha | Fish | Fish |
| **21** | pussycat | Mammal | Mammal |
| **22** | seahorse | Fish | Fish |
| **23** | skimmer | Bird | Bird |
| **24** | slowworm | Reptiles | Reptiles |
| **25** | squirrel | Mammal | Mammal |
| **26** | toad | Frog | Frog |
| **27** | vampire | Mammal | Mammal |
| **28** | vole | Mammal | Mammal |
| **29** | vulture | Bird | Bird |
| **30** | wallaby | Mammal | Mammal |
| **31** | wolf | Mammal | Mammal |

Showing 1 to 11 of 31 entries, 3 total columns

**# take k=25**

knn\_zoo\_pred25<-knn(train = zoo\_train,test = zoo\_test,cl = zoo\_train\_labels, k=25)

knn\_zoo\_pred25

[1] Mammal Mammal Fish Fish Mammal Bird Mammal Fish Mammal Bird Marine Mammal

[13] Marine Marine Mammal Mammal Fish Mammal Bird Fish Mammal Bird Bird Bird

[25] Mammal Bird Mammal Mammal Bird Mammal Mammal

Levels: Mammal Bird Reptiles Fish Frog Insects Marine

**#calculate the proportion of correct classification for k=1**

zoo\_accur<-100\*sum(zoo\_test\_labels==knn\_zoo\_pred25)/NROW(zoo\_test\_labels)

zoo\_accur

77.41935

View(data.frame( zoo\_animals,zoo\_test\_labels,knn\_zoo\_pred25))

|  | **zoo\_animals** | **zoo\_test\_labels** | **knn\_zoo\_pred25** |
| --- | --- | --- | --- |
|  |  |  |  |
| **1** | boar | Mammal | Mammal |
| **2** | calf | Mammal | Mammal |
| **3** | carp | Fish | Fish |
| **4** | catfish | Fish | Fish |
| **5** | cheetah | Mammal | Mammal |
| **6** | crow | Bird | Bird |
| **7** | deer | Mammal | Mammal |
| **8** | dogfish | Fish | Fish |
| **9** | dolphin | Mammal | Mammal |
| **10** | duck | Bird | Bird |
| **11** | gnat | Insects | Marine |
| **12** | hamster | Mammal | Mammal |
| **13** | honeybee | Insects | Marine |
| **14** | ladybird | Insects | Marine |
| **15** | lion | Mammal | Mammal |
| **16** | mongoose | Mammal | Mammal |
| **17** | newt | Frog | Fish |
| **18** | opossum | Mammal | Mammal |
| **19** | parakeet | Bird | Bird |
| **20** | piranha | Fish | Fish |
| **21** | pussycat | Mammal | Mammal |
| **22** | seahorse | Fish | Bird |
| **23** | skimmer | Bird | Bird |
| **24** | slowworm | Reptiles | Bird |
| **25** | squirrel | Mammal | Mammal |
| **26** | toad | Frog | Bird |
| **27** | vampire | Mammal | Mammal |
| **28** | vole | Mammal | Mammal |
| **29** | vulture | Bird | Bird |
| **30** | wallaby | Mammal | Mammal |
| **31** | wolf | Mammal | Mammal |

Showing 1 to 11 of 31 entries, 3 total columns