

TABLE OF CONTENTS

Question : Prepare a model for glass classification using KNN	1
Data Pre Processing	1
Summary of Data	1
Proportion for Glass Type	2
User Function:	2
Model Prediction	2
Model 1 (K = 5)	2
MODEL 2 (K = 5)	2
Model (K = ALL odds from 1 to 50)	3
Final Model (k=3)	3
Question 2 : Implement a KNN model to classify the animals in to categories	4
Model PREDICTION	4
Model 1 (k = 3)	4
Model k= all odds from 1 to 50	5
Model k = 1	5

QUESTION NO 1

QUESTION : PREPARE A MODEL FOR GLASS CLASSIFICATION USING KNN

In our data our Target variable is Type, i.e. representing the type of the glass.

DATA PRE PROCESSING

SUMMARY OF DATA

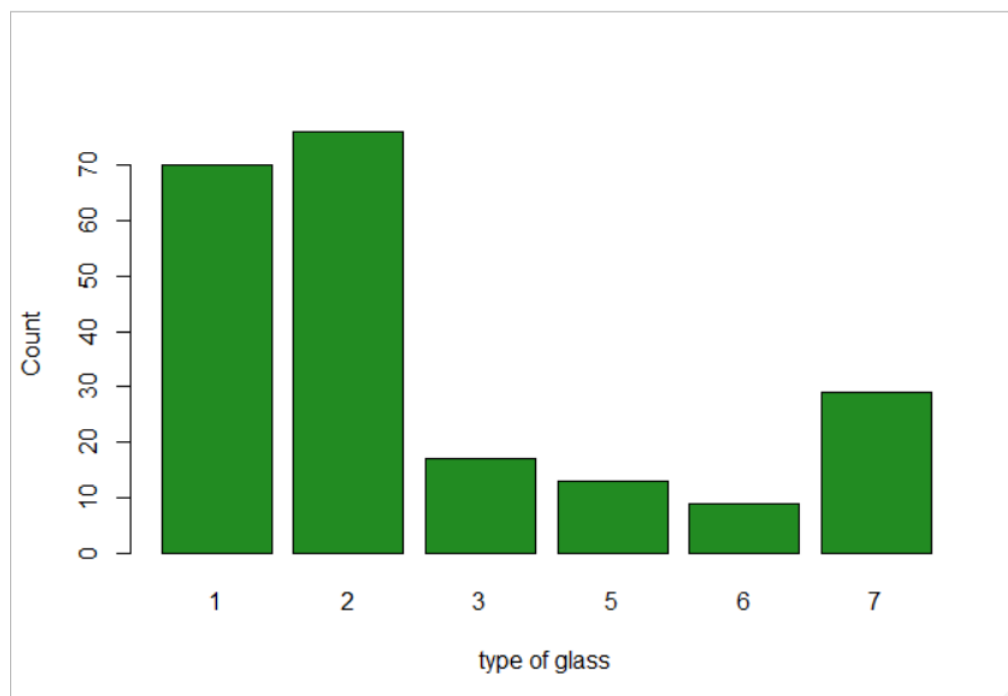
	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe
<i>Min</i>	1.511	10.73	0.000	0.290	69.81	0.0000	5.430	0.000	0.00000
<i>1st Qu.</i>	1.517	12.91	2.115	1.190	72.28	0.1225	8.240	0.000	0.00000
<i>Median</i>	1.518	13.30	3.480	1.360	72.79	0.5550	8.600	0.000	0.00000
<i>Mean</i>	1.518	13.41	2.685	1.445	72.65	0.4971	8.957	0.175	0.05701
<i>3rd Qu.</i>	1.519	13.82	3.600	1.630	73.09	0.6100	9.172	0.000	0.10000
<i>Max.</i>	1.534	17.38	4.490	3.500	75.41	6.2100	16.190	3.150	0.51000

Here We can see that chances of getting outlier in our data is very rare, cause mean and median is approximately same for every variable.

PROPERTIONFOR GLASS TYPE

1	2	3	5	6	7
32.71	35.51	7.94	6.07	4.21	13.55

The column is somewhat balanced and we may rely on this target variable.



USER FUNCTION:

Here I am using the same function, which was created by me during my Clustering Assignment, [normalize_dummy](#)

I used the above function to normalize my data

MODEL PREDICTION

MODEL 1 (K = 5)

In my model 1 I use the function knn3 from the library caret, and come up with accuracy of 0.6491. With confusion matrix as given below. Here I have considered my K as 5.

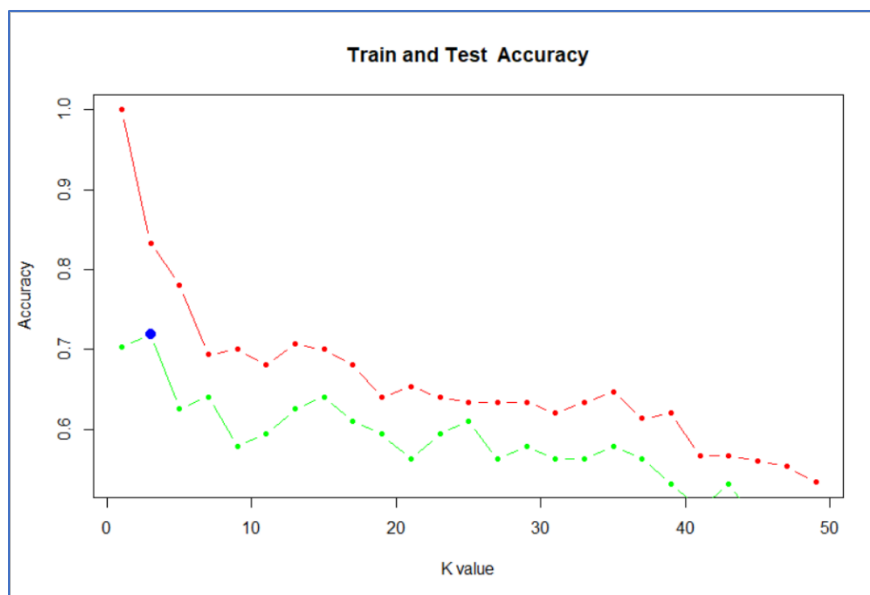
```
pred_1
  1  2  5  6  7
1 13  7  0  0  0
2  5 13  1  0  0
3  3  0  0  0  0
5  0  0  1  0  0
6  0  1  0  1  0
7  1  1  0  1  9
```

MODEL 2 (K = 5)

Then I do my second model using the function knn from the library class. Here I come up with accuracy of 0.625, with confusion matrix given below.

```
model_2
      1  2  3  5  6  7
1  14  7  0  0  0  0
2   5 14  0  1  1  0
3   3  1  1  0  0  0
5   0  2  0  1  0  0
6   0  1  0  0  1  0
7   1  1  0  0  1  9
```

MODEL (K = ALL ODDS FROM 1 TO 50)



Here We have two lines, one is red representing the Train accuracy, another is green representing the test accuracy.

Here We can see that train accuracy is very high for value 1 and test accuracy is high for k=3.

Now I tried to improve my accuracy, to do that I construct model for all k odd values from 1 to 50 and calculated the accuracy for test data and train data, after getting accuracies for each and every point, I plotted it .

Here from the graph I came to know that for k as 3 we are getting the highest accuracy in our model.

FINAL MODEL (K=3)

I build the model with considering my k as 3 and come up with accuracy as 0.718575.

```
model_final
      1  2  3  5  6  7
1  16  5  0  0  0  0
2   5 15  0  1  0  0
3   3  1  1  0  0  0
5   0  0  0  3  0  0
6   0  0  0  0  2  0
7   1  1  0  0  1  9
```

The above table is representing the confusion matrix for our final model.

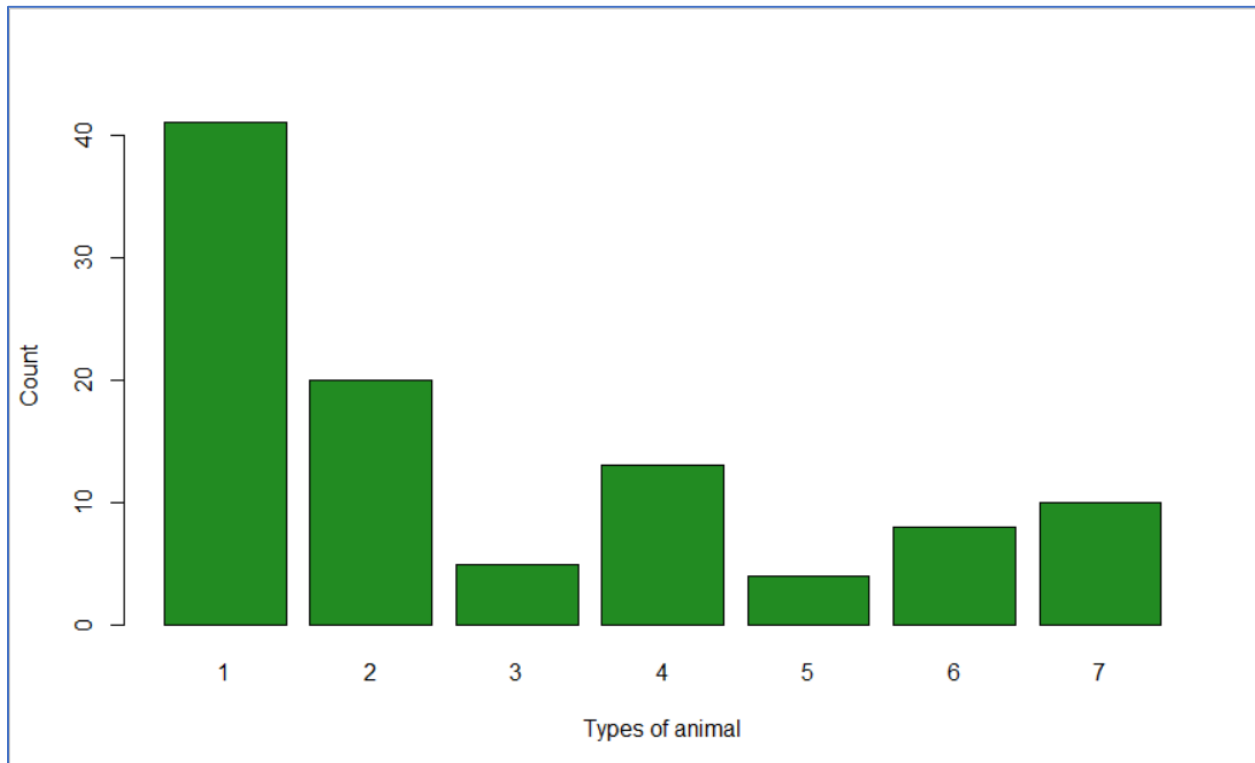
QUESTION NO 2

QUESTION 2 : IMPLEMENT A KNN MODEL TO CLASSIFY THE ANIMALS INTO CATEGORIES

Here Our target variable is Type of animals

```
1 2 3 4 5 6 7
41 20 5 13 4 8 10
```

There are 7 types of animal in our zoo



From the above plot we can see that the animal type 1 is highest among all.

MODEL PREDICTION

MODEL 1 (K = 3)

In my first model I considered the k as 3 and come up with 0.9 accuracy. with the confusion matrix given below.

```
1 2 3 4 5 6 7
1 10 0 0 0 0 0 0
2 0 7 0 0 0 0 0
3 0 0 0 1 1 0 0
4 0 0 0 5 0 0 0
5 0 0 0 0 1 0 0
6 0 0 0 0 0 3 0
7 0 0 0 0 0 1 1
```

Here I come up with 0.9 accuracy.

Now I move to another model for higher accuracy if available.

MODEL K= ALL ODDS FROM 1 TO 50

Here I considered all the possible Train accuracies and test accuracies for my model. Then I plotted it.



From the plot we can see that for k=1 we are getting highest test accuracies.

MODEL K = 1

In my final model I assume k as 1 for getting the highest accuracy. Now I come up with confusion matrix

zoomodel	1	2	3	4	5	6	7
1	10	0	0	0	0	0	0
2	0	7	0	0	0	0	0
3	0	0	2	0	0	0	0
4	0	0	0	5	0	0	0
5	0	0	0	0	1	0	0
6	0	0	0	0	0	3	0
7	0	0	1	0	0	0	1

Here I come up with 0.966667

Only one type of animal 7 is miss classified as type 3. So, I can rely in this model.