NAME: Penugonda Abhignanendra   **REGNO**:19BCE1677

FACULTY: Dr. Lakshmi Pathi Jakkamputi   **SLOT**: L21, L22

Essential of Data Analytics
Tasks for Week-6: K-NN Algorithm

Aim: Understand the following operations/functions on to perform K- NN algorithm and perform similar operations on 'wdbc' dataset based on given instructions.

Algorithm:

- removing all the values from the global environment.
- As our dataset is not csv so we are using file.choose() to choose our dataset and use it for prediction.
- To see the dataset use view() function.
- Mynorm function is created to find the normalize the values separating each and every value with its min value and dividing it the difference of max and min value in the column.
- Create a new dataframe named mydata and store all the normalized value of every in that new dataframe except the first column as it is an categorical data.
- For comparing the original dataset and normalized dataset take 2 to 5 columns of both data set and apply summary() function to find summary.
- Divide the first 400 values as the train dataset and remaining 169 values to the test dataset from mydata(normalized dataset).
- Import class using library() function.
- Do the knn algorithm and store all the predicted values in pred variable.
- Form the confusion matrix form using predicted data from pred variable and form 401 to 569 rows in first dataset.
- Find the accuracy of the data by adding the [1,1] element and [2,2] element and dividing its summation with the whole sum.

Inference:

The accuracy of the model is 97%. so, we can say that the model is best fit model.

Result:

Essential of Data Analytics
Tasks for Week-6: K-NN Algorithm

Confusion matrix:

pred

pred	B	M
B	128	3
M	2	36

Accuracy:

0.9704142

Code:

```
rm(list=ls())

wdbc<-read.table(file.choose(),sep=',')

view(wdbc)

wdbc<-wdbc[,-1]

mynorm<-function(x){((x-min(x))/(max(x)-min(x)))}

mydata<-as.data.frame(lapply(wdbc[,-1], mynorm))

summary(wdbc[,2:5])

summary(mydata[,1:4])

train<-mydata[1:400,]

test<-mydata[401:569,]

library(class)

pred<-knn(train,test,wdbc[1:400,1],k=21)

cf<-table(pred,wdbc[401:569,1])

cf

acc=(cf[[1,1]]+cf[[2,2]])/sum(cf)

acc
```