	Marwadi University Faculty of Technology Department of Information and Communication Technology
Subject: Machine Learning (01CT0607)	**Aim: To obtain the classification of various classes using k-Nearest Neighbor approach**
Experiment No: 07	**Date:07-09-2022**	**Enrollment No:92000133018**

Aim: To obtain the classification of various classes using k-Nearest Neighbor approach

IDE: Google Colab

Theory:

This algorithm is used to solve the classification model problems. K-nearest neighbor or K-NN algorithm basically creates an imaginary boundary to classify the data. When new data points come in, the algorithm will try to predict that to the nearest of the boundary line. Therefore, larger k value means smother curves of separation resulting in less complex models. Whereas, smaller k value tends to over fit the data and resulting in complex models. It's very important to have the right k-value when analyzing the dataset to avoid over fitting and under fitting of the dataset.

The model representation for KNN is the entire training dataset. It is as simple as that. KNN has no model other than storing the entire dataset, so there is no learning required. Efficient implementations can store the data using complex data structures like k-d trees to make look-up and matching of new patterns during prediction efficient. Because the entire training dataset is stored, you may want to think carefully about the consistency of your training data. It might be a good idea to curate it, update it often as new data becomes available and remove erroneous and outlier data.

K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.

K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

Example: Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.

	Marwadi University
![Marwadi University Logo]	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**
Subject: Machine Learning (01CT0607)	**Aim: To obtain the classification of various classes using k-Nearest Neighbor approach**
Experiment No: 07	**Date:07-09-2022**	**Enrollment No:92000133018**

Methodology:

1. Load the basic libraries and packages
2. Load the dataset
3. Analyse the dataset
4. Pre-process the data

5. Visualize the Data
6. Separate the feature and prediction value columns
7. Select the number K of the neighbors
8. Calculate the Euclidean distance of K number of neighbors
9. Take the K nearest neighbors as per the calculated Euclidean distance.
10. Among these k neighbors, count the number of the data points in each category.
11. Assign the new data points to that category for which the number of the neighbor is maximum.
12. Our model is ready.

	Marwadi University **Faculty of Technology** **Department of Information and Communication Technology**
Subject: Machine Learning (01CT0607)	**Aim: To obtain the classification of various classes using k-Nearest Neighbor approach**
Experiment No: 07	**Date:07-09-2022** **Enrollment No:92000133018**

Program (Code):

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import math

# loading the data
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
col_names = ["sepal length","sepal width", "petal length", "petal width", "class"]
dataset = pd.read_csv(url,names=col_names)

dataset

dataset.shape

training = dataset.values[:,0:4]

training

trainingclass = dataset.values[:,-1]
trainingclass

def euclidean_distance(A,B):
  distance =0
  for i in range(len(A)):
    distance+=(A[i]-B[i])**2
  distance = math.sqrt(distance)
  return distance

unique_class = list(set(trainingclass))

unique_class


#converting label to numerical labels

for i in range(len(trainingclass)):
  for k in range(len(unique_class)):
    if(trainingclass[i]==unique_class[k]):
      trainingclass[i]=k
print(trainingclass)
```

	Marwadi University
![Marwadi University logo]	**Faculty of Technology**
	Department of Information and Communication Technology
Subject: Machine Learning (01CT0607)	**Aim: To obtain the classification of various classes using k-Nearest Neighbor approach**
Experiment No: 07	**Date:07-09-2022**	**Enrollment No:92000133018**

```
testing= training[0]
testing

distance = []
for i in range(len(training)):
  distance.append([euclidean_distance(training[i],testing),trainingclass[i]])

distance

distance.sort()
distance

k=5

#distance[:k+1]
identify = [0 for i in range(len(unique_class))]

identify

for i in range(1,k+1):
  identify[distance[i][1]]+=1
print(identify)

print("the predicted class is : ",unique_class[identify.index(max(identify))])
```

Results:

To be attached with

    a.   Confusion matrix

![Marwadi University logo] Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**
Subject: Machine Learning (01CT0607)	**Aim: To obtain the classification of various classes using k-Nearest Neighbor approach**
Experiment No: 07	**Date:07-09-2022**	**Enrollment No:92000133018**

```
[38] k=5
```

```
[39] #distance[:k+1]
     identify = [0 for i in range(len(unique_class))]
```

```
[42] identify

     [0, 0, 5]
```

```
for i in range(1,k+1):
  identify[distance[i][1]]+=1
print(identify)
```

```
[0, 0, 5]
```

b. Accuracy Score

```
[41] print("the predicted class is : ",unique_class[identify.index(max(identify))])

     the predicted class is :  Iris-setosa
```

Observation and Result Analysis:

a. Nature of the dataset

_____

_____

_____

_____

b. During Training Process

_____

_____

_____

_____

c. After the training Process

	Marwadi University
![Marwadi University logo] **Marwadi University**	**Faculty of Technology**
	Department of Information and Communication Technology
Subject: Machine Learning (01CT0607)	**Aim: To obtain the classification of various classes using k-Nearest Neighbor approach**
Experiment No: 07	**Date:07-09-2022**	**Enrollment No:92000133018**

_____

_____

_____

_____

d. Observation over the Learning Curve

_____

_____

_____

Post Lab Exercise:

a. Is KNN useful in regression-based problems?

_____

_____

_____

b. Why KNN is a non-parametric algorithm?

_____

_____

_____

c. Which are the types of distance metrices?

_____

_____

_____

d. What can be the criteria of selection of the value of K?

_____

_____

_____

![Marwadi University logo]	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0607)	**Aim: To obtain the classification of various classes using k-Nearest Neighbor approach**	
Experiment No: 07	**Date:07-09-2022**	**Enrollment No:92000133018**

e. What are the assumptions of KNN approach?

_____

_____

_____

f. How can the KNN approach make the prediction of the unseen dataset?

_____

_____

_____

g. Why is KNN called a Lazy Learner?

_____

_____

_____

h. Why should we not use KNN for large dataset?

_____

_____

_____

i. What are the advantages of KNN algorithm?

_____

_____

_____

Post Lab Activity:

Consider any dataset from **https://archive.ics.uci.edu/ml/datasets.php** and perform the classification using KNN algorithm. Make sure that the dataset is not matching with your classmates. You can also select the dataset from other ML repositories with prior permission from your concerned subject faculty.