Marwadi University	Marwadi University Faculty of Technology Department of Information and Communication Technology	
Subject: Artificial Intelligence (01CT0616)	Aim: To obtain the classification of various classes using k-Nearest Neighbor approach	
Experiment No: 03	Date:	Enrollment No: 92200133030

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.

Methodology:

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

Marwadi University	Marwadi University Faculty of Technology Department of Information and Communication Technology	
Subject: Artificial Intelligence (01CT0616)	Aim: To obtain the classification of various classes using k-Nearest Neighbor approach	
Experiment No: 03	Date:	Enrollment No: 92200133030

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.

Pre Lab Exercise:

a. What is K in KNN-approach?

In KNN, K is a hyper parameter. in KNN the distance of test data point is measured with the previously seen datapoint then top K will be selected.

b. Which are the types of distance metrices?

There are several types of distance metrics like
1) Euclidean Distance 2) Manhattan Distance
3) Minkowski Distance

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

We can use elbow method for selecting the K for are KNN model, it will identifying the point where adding more neighbors no longer

d. What are the advantages of KNN algorithm?

The advantages of KNN tends to 1) Simple and Easy 2) No Training phase 3) Handles Multiclass Problem 4) Flexible Distance Metrics.

Marwadi University	Marwadi University Faculty of Technology Department of Information and Communication Technology	
Subject: Artificial Intelligence (01CT0616)	Aim: To obtain the classification of various classes using k-Nearest Neighbor approach	
Experiment No: 03	Date:	Enrollment No: 92200133030

Observation and Result Analysis:

a. Nature of the dataset

The Dataset is having 4 colums sepal_length. Sepal_width, petal_Height, Petal_width with numerical values and 1 colums which is having the category of the flower.

b. During Training Process

In the Training Process, first of all KNN is not Having any training process that is why it is known as a lazy learners.

c. After the training Process

After the training process, model is capable to identify the category of the flower based on the distance of the input values.

d. Observation over the Learning Curve

observation over the learning curve is that in each iteration the accuracy of the model is increasing.

Marwadi University	Marwadi University Faculty of Technology Department of Information and Communication Technology	
Subject: Artificial Intelligence (01CT0616)	Aim: To obtain the classification of various classes using k-Nearest Neighbor approach	
Experiment No: 03	Date:	Enrollment No: 92200133030

Post Lab Exercise:

a. Is KNN useful in regression-based problems?

Yes, KNN can be used in the regression based problem by averaging the values of the k nearest neighbors.

b. Why KNN is a non-parametric algorithm?

It makes no assumptions about the underlying data distribution. it stores the entire dataset insted of learning a function during training.

c. What are the assumptions of KNN approach?

The Assumptions of KNN approachs are 1) Similarity principle, 2) Balanced data distribution. 3) Independent and Identically distributed

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

The calculations of the KNN includes 1) Distance calculation, 2) Find k neighbors 3) For classification it uses majority.

e. Why is KNN called a Lazy Learner?

KNN is called Lazy learner because it is not having any training process.

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

The KNN is not used for large dataset is because it is having high computational cost, it is memory Intensive.

10