# Iris Dataset Using KNN

PROJECT REPORT SUBMITTED AS A MANDATORY REQUIREMENT FOR THE DEGREE OF MASTER OF COMPUTER APPLICATION IN 2020-2022

SWAMI VIVEKANANDA UNIVERSITY

TELINIPARA, BARASAT - BARRACKPORE ROAD, BARA KANTHALIA, WEST BENGAL - 700121

**BY**

**AISWARYA ROUT**

**SEMESTER: 4th**

**REGISTRATION NUMBER**: **011-122-2020-001**

**DEPARTMENT OF MASTER OF COMPUTER APPLICATION SCHOOL OF COMPUTER SCIENCE**

### 2022

Swami Vivekananda University

Telinipara, Barasat - Barrackpore Road, Bara Kanthalia, West Bengal - 700121

**Department of Master of Computer Application School of Computer Science**

#### CERTIFICATE FROM THE SUPERVISOR

This is to certify that the project entitled “**Iris Dataset Using KNN**” has been prepared by **Ms. Aiswarya Rout** for the award of Degree of Master of Computer Application in Department of Master of Computer Application, Swami Vivekananda University, Barrackpore.

The research/survey work has been carried out under my supervision and the report is the result of a bonafide work. No part of the report has been submitted to this or to any other university for any degree. I recommend that the report may be placed before the examiners for consideration of award of the degree of this university.

(Signature of Supervisor)

(Date with official seal)

2

### DECLARATION

I, **Aiswarya Rout**, Registration Number **011-122-2020-001** hereby declare that this project entitled **‘Iris Dataset Using KNN**’ is submitted by me to the Swami Vivekananda University, Barrackpore, West Bengal for the degree of MCA. This bonafide work is completely done by me under the supervision of **Mrs. Sraddha Roy Choudhury**, Department of **Master of Computer Application**, under the School of **Computer Science** during **2020 – 2022**.

I, further declare that the results presented here have not been submitted before in any University for any degree or diploma.

Date: Place:

(Name of the Student)

3

**CONTENTS**

**TOPICS PAGE NO.**

**1.1 ABSTRACT 05**

**1.2 INTRODUCTION 06**

**1.3 DATASET 07**

**FIG.1. 07**

**FIG.2. 08**

**FIG.3. 08**

**1.4 ADVANTAGES & DISADVANTAGES 09**

**FIG.4 10**

**1.5 DATA ANALYSIS 11-24**

**1.6 CONCLUSION 25**

**1.7 REFERENCES 26**

4

**ABSTRACT**

K Nearest Neighbor is one of the fundamental algorithms in machine learning. Machine learning models use a set of input values to predict output values. KNN is one of the simplest forms of machine learning algorithms mostly used for classification. It classifies the data point on how its neighbor is classified. There are three types of machine learning are called Supervised, Unsupervised and Reinforcement learning. The Iris dataset and used K-Nearesr Neighbours (KNN) classification Algorithm. Iris dataset tools used Numpy, Pandas, Matplotlib and machine learning library Scikit-learn.

5

**INTRODUCTION**

Machine learning is about prediction on unseen data or testing data. In machine learning a computer first learn to perform a task by training dataset. Supervised learning is of two types Classification based and Regression based. KNN is a simple algorithm that stores all available cases and classifies based on a similarity measures.

The implementation of the model includes six basic steps of machine learning that are:

1. Collect data/prepare data
2. Choose algorithm
3. Creating object of the model
4. Train the model by training dataset
5. Making prediction on unseen data or testing data
6. Evaluation of the model

The dataset contain 150 Samples of data has 3 classes, each contain 50 smples. To train the machine we split the dataset into two Parts training and testing dataset and then it will test on testing dataset.

6

**DATASET**

The Iris dataset from Scikit learn in which iris dataset is already inbuilt. The dataset contain 150 sample data in it. The dataset has three classes of data that are Setosa, Versicolor and Virginica each having 50 sample data. Number of attributes in the datasets are :

4 numeric attributes, predictive attribute (class of iris plant) and the class attribute information.

1. Sepal length in cm
2. Sepal width in cm
3. Ppetal length in cm
4. Petal width in cm



**Fig.1. Iris Setosa**

There are three types of

machine learning are called as Supervised, Unsupervised and

Reinforcement learning. In this paper we have worked on

supervised learning. We have taken the iris dataset and used

K-Nearest Neighbors (KNN) classification Algorithm. Our

purpose is build the model that is able to automatically

recognize the iris species. Tools used for this in paper are

Numpy, Pandas, Matplotlib and machine learning library

Scikit-learn.

There are three types of

machine learning are called as Supervised, Unsupervised and

Reinforcement learning. In this paper we have worked on

supervised learning. We have taken the iris dataset and used

K-Nearest Neighbors (KNN) classification Algorithm. Our

purpose is build the model that is able to automatically

recognize the iris species. Tools used for this in paper are

Numpy, Pandas, Matplotlib and machine learning library

Scikit-learn.

There are three types of

machine learning are called as Supervised, Unsupervised and

Reinforcement learning. In this paper we have worked on

supervised learning. We have taken the iris dataset and used

K-Nearest Neighbors (KNN) classification Algorithm. Our

purpose is build the model that is able to automatically

recognize the iris species. Tools used for this in paper are

Numpy, Pandas, Matplotlib and machine learning library

Scikit-learn.

There are three types of

machine learning are called as Supervised, Unsupervised and

Reinforcement learning. In this paper we have worked on

supervised learning. We have taken the iris dataset and used

K-Nearest Neighbors (KNN) classification Algorithm. Our

purpose is build the model that is able to automatically

recognize the iris species. Tools used for this in paper are

Numpy, Pandas, Matplotlib and machine learning library

Scikit-learn.

There are three types of

machine learning are called as Supervised, Unsupervised and

Reinforcement learning. In this paper we have worked on

supervised learning. We have taken the iris dataset and used

K-Nearest Neighbors (KNN) classification Algorithm. Our

purpose is build the model that is able to automatically

recognize the iris species. Tools used for this in paper are

Numpy, Pandas, Matplotlib and machine learning library

Scikit-learn.

7



**Fig.2. Iris Versicolor**

****

**Fig.3.Iris Virginica**

8

**ADVANTAGES & DISADVATAGES IRIS DATASET**

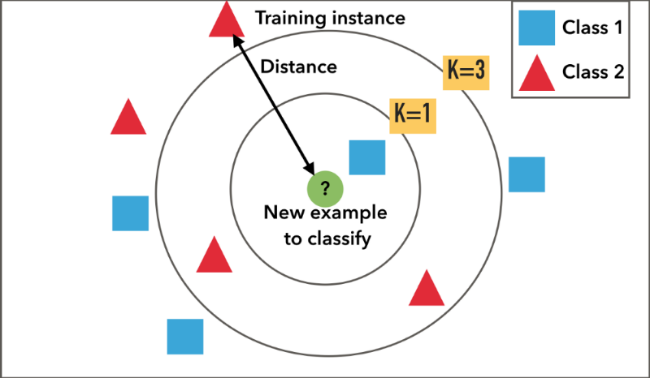
**Advantages:**

1. No Training Period- KNN modeling does not include training period as the data itself is a model which will be the reference for future prediction and because of this it is very time efficient in term of improvising for a random modeling on the available data.
2. Easy Implementation- KNN is very easy to implement as the only thing to be calculated is the distance between different points on the basic of data of different features and this distance can easily be calculated using distance formula sach as Euclidian or Manhattan
3. No Training Time for classification/regression- The KNN algorithm has no explicit training step and all the work happens during prediction.

**Disadvantages:**

1. Does not work well with large dataset as calculating distances between each data instance would be very costly.
2. Does not work with high dimensionality as this will complicate the distance calculating process to calculate distance for each dimension.
3. Sensitive to noisy and missing data
4. Feature Scaling Data in all the dimension should be scaled properly.

9



**Fig.4. Advantages and disadvantages KNN**

10

**CONCLUSION**

**KNN Conceptually simple, easy to understand and explain. It is Very flexible decision boundaries. Not much learning at all. It can be hard to find a good distance measure. Irrelevant features and noise can be very detrimental. Typically cannot handle more than a few dozen attributes. Computational cost requires a lot computation.**

11

**REFERENCES**

1: <https://machinelearningmastery.com/tutorial-to-implement-k-nearest-neighbors-in-python-from-scratch/>

2: <https://www.kaggle.com/arshid/iris-flower-dataset>

3: [https://www.bing.com/search?q=iris+dataset+references&qs=n&form=QBRE&sp=-1&pq=iris+dataset+references&sc=8-23&sk=&cvid=5CDC8CBE06334C958734832B1705F6E5#](https://www.bing.com/search?q=iris+dataset+references&qs=n&form=QBRE&sp=-1&pq=iris+dataset+references&sc=8-23&sk=&cvid=5CDC8CBE06334C958734832B1705F6E5)

4: <https://medium.com/.../exploratory-data-analysis-iris-dataset-4df6f045cda>

5: [https://www.bing.com/search?q=iris+dataset+references&qs=n&form=QBRE&sp=-1&pq=iris+dataset+references&sc=8-23&sk=&cvid=5CDC8CBE06334C958734832B1705F6E5#](https://www.bing.com/search?q=iris+dataset+references&qs=n&form=QBRE&sp=-1&pq=iris+dataset+references&sc=8-23&sk=&cvid=5CDC8CBE06334C958734832B1705F6E5)

6: <https://arcca.github.io/.../03-scikit-learn-iris-dataset/index.html>

7: <https://www.geeksforgeeks.org/exploratory-data-analysis-on-iris-dataset>

8: <https://machinelearningmastery.com/tutorial-to-implement-k-nearest-neighbors-in-python-from-scratch/>

9: <https://valueml.com/understanding-knn-algorithm-using-iris-dataset-with-python>

10: <https://kirenz.github.io/classification/docs/knn-iris.html>

12