

# Decision Tree

CSE-0408 Summer 2021

Md. Mazharul Haque

Department of Computer Science and Engineering

State University of Bangladesh (SUB)

Dhaka, Bangladesh

mazharulsub47@gmail.com

**Abstract**—Decision tree classifiers are widely recognized as one of the most well-known approaches for representing data classification in classifiers. The topic of extending a decision tree using existing data has been studied by researchers from diverse domains and backgrounds, including machine learning, pattern recognition, and statistics. Decision tree classifiers have been proposed in a variety of disciplines, including medical disease analysis, text categorization, user smartphone classification, pictures, and many others.

**Index Terms**—Artificial Intelligence, Machine Learning, Supervised, Classification, Decision Tree.

## I. INTRODUCTION

Classification is a two-step process, learning step and prediction step, in machine learning. In the learning step, the model is developed based on given training data. In the prediction step, the model is used to predict the response for given data. Decision Tree is one of the easiest and popular classification algorithms to understand and interpret.

## II. LITERATURE REVIEW

Assegie and Nair used the DT classification technique to categorize the handwritten digits in the kaggle digits standard data set and assess the model's accuracy for each digit from 0 to 9. The kaggle features comprise 42,000 rows and 720 columns for machine learning, as well as vector characteristics for digital image pixels. They applied machine learning algorithms to map the classifier's success rate graph in the reality of handwritten digits using a highly efficient language called "python programming." The 83.4 percent accuracy and decision tree classifier had an impact on handwritten number recognition, according to the findings.

## III. DECISION TREE ALGORITHM

Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too.

The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data(training data).

In Decision Trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of

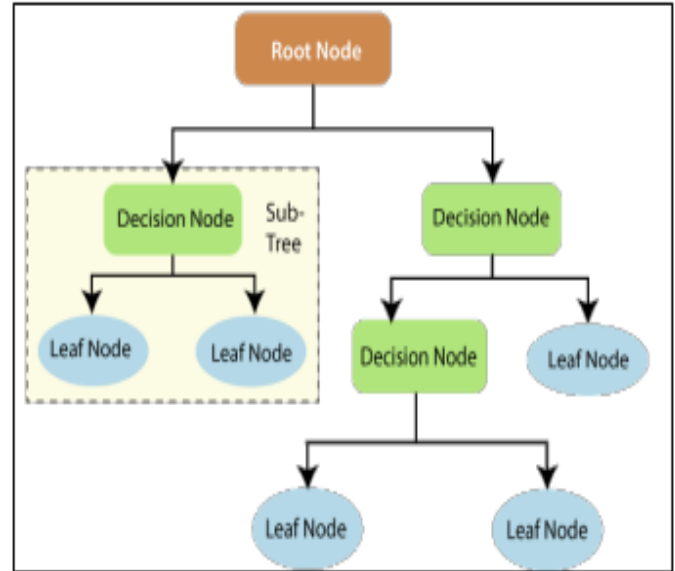


Fig. 1. Decision Tree

comparison, we follow the branch corresponding to that value and jump to the next node.

### Libraries Requirements

- *pandas*
- *sklearn*
- *IPython*
- *matplotlib*

## IV. CONCLUSION

This assignment is based on a graphic representation of a decision tree. A data-set is given for the training and visualization of this decision tree.

### ACKNOWLEDGMENT

I would like to thank my honourable **Khan Md. Hasib Sir** for his time, generosity and critical insights into this project.

### REFERENCES

- [1] D. Abdulqader, A. Mohsin Abdulazeez, and D. Zeebaree, "Machine Learning Supervised Algorithms of Gene Selection: A Review," Apr. 2020.

# K- Nearest Neighbors

CSE-0408 Summer 2021

Md. Mazharul Haque

*Department of Computer Science and Engineering*

*State University of Bangladesh (SUB)*

Dhaka, Bangladesh

mazharulsub47@gmail.com

**Abstract**—we are going to implement K-Nearest Neighbors using Jupyter Notebook.

**Index Terms**—K nearest neighbors.

## REFERENCES

- [1] Solichin, A. (2019, September). Comparison of Decision Tree, Naïve Bayes and K-Nearest Neighbors for Predicting Thesis Graduation. In 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI) (pp. 217-222). IEEE.

## I. INTRODUCTION

we will implement another widely used machine learning classification technique called K-nearest neighbors (KNN). Our focus will be primarily on how does the algorithm work and how does the input parameter affects the output/prediction.

## II. KNN ALGORITHM

We can implement a KNN model by following the below steps:

- *Step-1: Select the number  $K$  of the neighbors*
- *Step-2: Calculate the Euclidean distance of  $K$  number of neighbors*
- *Step-3: Take the  $K$  nearest neighbors as per the calculated Euclidean distance.*
- *Step-4: Among these  $k$  neighbors, count the number of the data points in each category.*
- *Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.*
- *Step-6: Our model is ready.*

## Libraries Requirements

- *pandas*
- *sklearn*
- *matplotlib*

**Pandas** is used to take input data sets, **sklearn** is used to develop and train our models, as well as **matplotlib** are used to visualize our K-Nearest Neighbors accuracy graphically.

## III. CONCLUSION

This assignment is based on a graphic representation of a KNN model accuracy. A data-set is given for the training and visualization of this KNN model accuracy.

## ACKNOWLEDGMENT

I would like to thank my honourable **Khan Md. Hasib Sir** for his time, generosity and critical insights into this project.