# Final Report-1
# Title:Decision Tree

CSE-0408 Summer 2021

Name:Farhan Mashuk Saumik
*Department of Computer Science and Engineering*
*State University of Bangladesh (SUB)*
Dhaka, Bangladesh
email:farhanmashuksaumik1996@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, Classificatio**

## I. Introduction

A decision tree is a flowchart-like tree structure where an internal node represents feature(or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree in recursively manner call recursive partitioning. This flowchart-like structure helps you in decision making. It's visualization like a flowchart diagram which easily mimics the human level thinking. That is why decision trees are easy 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

Step-1: Begin the tree with the root node, says S, which contains the complete dataset. Step-2: Find the best attribute in the dataset using Attribut Selection Measure (ASM). Step-3: Divide the S into subsets that contains possible values for the best attributes. Step-4: Generate the decision tree node, which contains the best attribute. Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

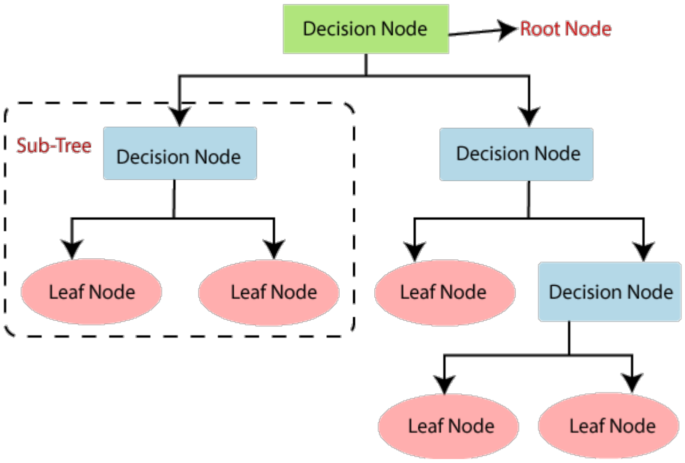

Fig. 1. Dicision Tree

## IV. DECISION TREE ADVANTAGES AND DISADVANTAGES

Advantages
– Does not require normalization of data.
– Does not require scaling of data as well.
Disadvantages
– Often involves higher time to train the model.
– Training is relatively expensive as the complexity and time has taken are more.

## V. 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] [1] D. Abdulqader, A. Mohsin Abdulazeez, and D. Zeebaree, “Machine Learning Supervised Algorithms of Gene Selection: A Review,” Apr. 2020

# Final Report-2
# Title:K-Nearest Neighbors (KNN)

CSE-0408 Summer 2021

Name:Farhan Mashuk Saumik
*Department of Computer Science and Engineering*
*State University of Bangladesh (SUB)*
Dhaka, Bangladesh
email:farhanmashuksaumik1996@gmail.com

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

***Index Terms*—K nearest neighbors.**

## I. Introduction

KNN is a non-parametric and lazy learning algorithm. Non-parametric means there is no assumption for underlying data distribution. In other words, the model structure determined from the dataset. This will be very helpful in practice where most of the real world datasets do not follow mathematical theoretical assumptions. Lazy algorithm means it does not need any training data points for model generation. All training data used in the testing phase. This makes training faster and testing phase slower and costlier. Costly testing phase means time and memory. In the worst case, KNN needs more time to scan all data points and scanning all data points will require more memory for storing training data.

## II. Literature Review

I am using python to solving this.

## III. 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.

## IV. KNN ADVANTAGES AND DISADVANTAGES

Advantages
- Quick calculation time.
- Simple algorithm – to interpret.
- Versatile – useful for regression and classification.

Disadvantages
- Does not work well with large dataset.
- Does not work well with high dimensions.

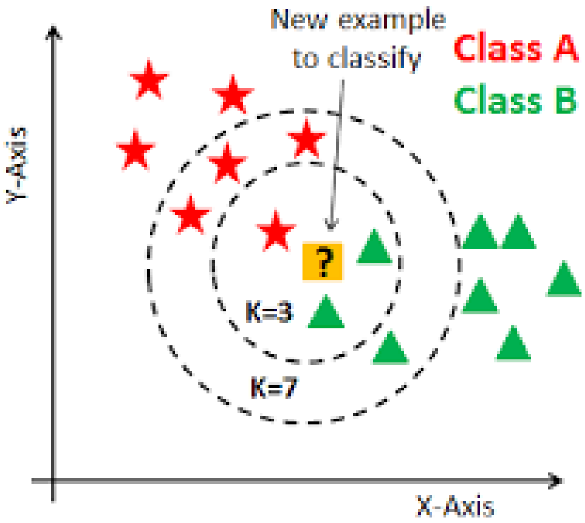

Fig. 1. How does the KNN algorithm work

## V. 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.

## References

[1] Solichin, A. (2019, September). Comparison of Decision Tree, Na¨ıve 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