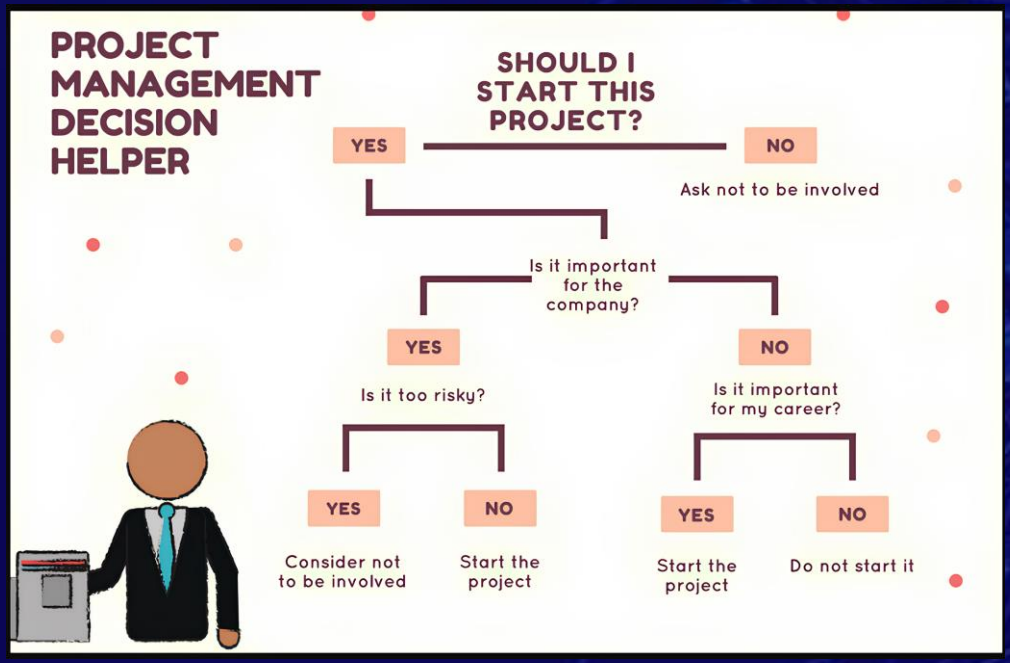
Unit 2.4
Decision Tree



Reference

Disclaimer

The content is curated from online/offline resources and used for educational purpose only

Decision Tree

Reference

Learning Objectives

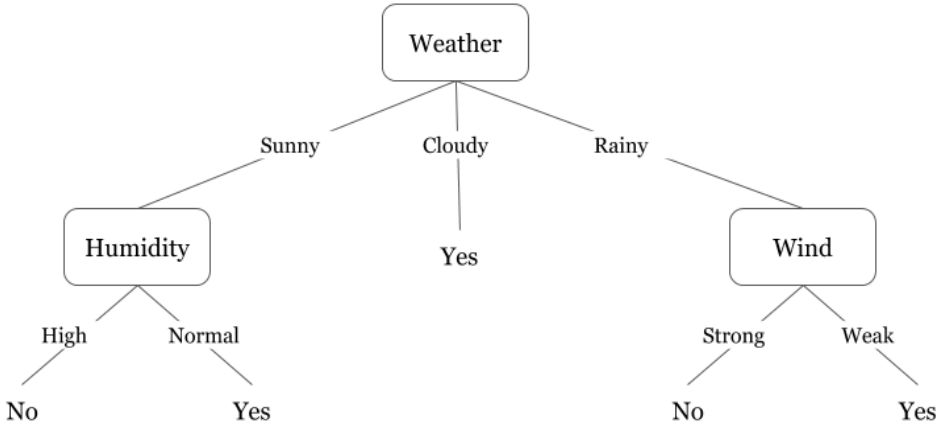
You will learn in this lesson:

- Concept of Decision Tree
- Use of Decision Tree to classify data
- Basic algorithm to build Decision Tree
- Some illustrations
- Concept of Entropy
- Basic concept of entropy in information theory
- Mathematical formulation of entropy
- Calculation of entropy of a training set
- Decision Tree induction algorithms
- ID3

Decision Tree

Introduction

- Basic idea behind building a decision tree is to map all the possible decision paths in the form of a tree.
- Efficient machine learning algorithm.
- Need to create new tree once seen whole new data
- Data driven programing the conditions.



Weather Prediction

Decision Tree Important Terminology

Root Node: Root node is from where the decision tree starts. It represents the entire dataset, which further gets divided into two or more homogeneous sets.

Leaf Node: Leaf nodes are the final output node, and the tree cannot be segregated further after getting a leaf node.

Splitting: Splitting is the process of dividing the decision node/root node into sub-nodes according to the given conditions.

Branch/Sub Tree: A tree formed by splitting the tree.

Pruning: Pruning is the process of removing the unwanted branches from the tree.

Parent/Child node: The root node of the tree is called the parent node, and other nodes are called the child nodes.

Case Study: Drug Prediction

- A medical researcher compiling data for a study.
- During course of treatment, each patient responded to one of two medications; we'll call them Drug A and Drug B.
- Job is to build a model to find out which drug might be appropriate for a patient with same illness.
- Feature sets: Age, Gender, Blood Pressure, and Cholesterol
- Target: Drug that each patient responded.

	Features				Target
Patient ID	**Age**	**Sex**	**BP**	**Cholesterol**	**Drug**
p1	Young	F	High	Normal	Drug A
p2	Young	F	High	High	Drug A
p3	Middle-age	F	Hiigh	Normal	Drug B
p4	Senior	F	Normal	Normal	Drug B
p5	Senior	M	Low	Normal	Drug B
p6	Senior	M	Low	High	Drug A
p7	Middle-age	M	Low	High	Drug B
p8	Young	F	Normal	Normal	Drug A
p9	Young	M	Low	Normal	Drug B
p10	Senior	M	Normal	Normal	Drug B
p11	Young	M	Normal	High	Drug B
p12	Middle-age	F	Normal	High	Drug B
p13	Middle-age	M	High	Normal	Drug B
p14	Senior	F	Normal	High	Drug A
p15	Middle-age	F	Low	Normal	?

Patient Drug Data

How does the Decision Tree algorithm Work?

The complete process can be better understood using the below 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 Attribute 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

How to Select the Effective Attribute ?

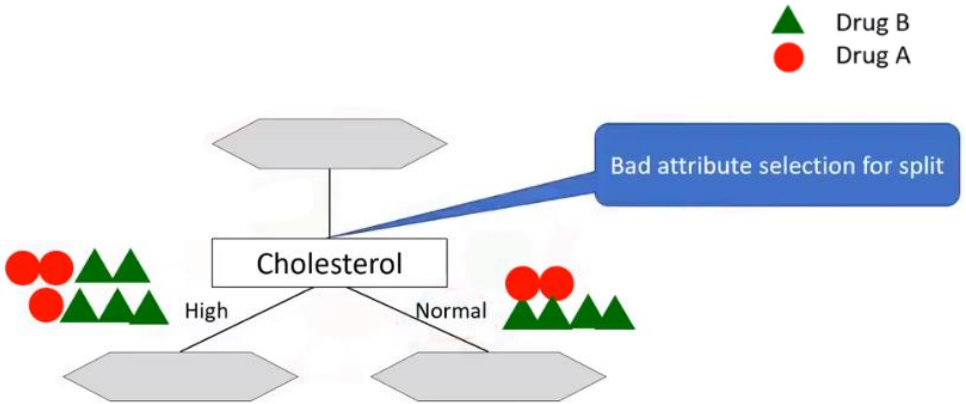
- The algorithm chooses the most predictive feature to split the data on.
- Root node can be any valid feature.
- Root node divides to create branches.

Patient ID	Age	Sex	BP	Cholesterol	Drug
p1	Young	F	High	Normal	Drug A
p2	Young	F	High	High	Drug A
p3	Middle-age	F	Hiigh	Normal	Drug B
p4	Senior	F	Normal	Normal	Drug B
p5	Senior	M	Low	Normal	Drug B
p6	Senior	M	Low	High	Drug A
p7	Middle-age	M	Low	High	Drug B
p8	Young	F	Normal	Normal	Drug A
p9	Young	M	Low	Normal	Drug B
p10	Senior	M	Normal	Normal	Drug B
p11	Young	M	Normal	High	Drug B
p12	Middle-age	F	Normal	High	Drug B
p13	Middle-age	M	High	Normal	Drug B
p14	Senior	F	Normal	High	Drug A
p15	Middle-age	F	Low	Normal	?

Patient Drug Data

Effective Attribute Quest!

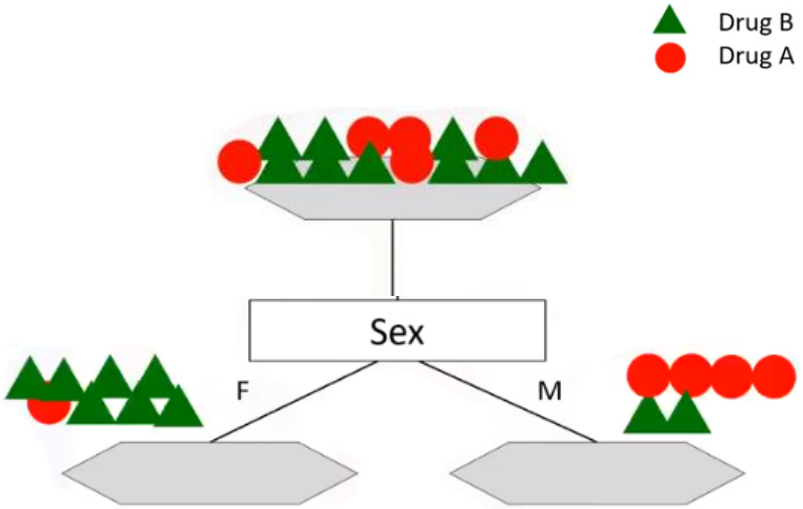
- Let’s begin our quest by picking “Cholesterol” as the first attribute to split data.
- It is a sample of bad attribute selection for splitting data.
- Bad in terms of impurity cleanliness.
- Improper class balance attained.



How Decision Tree Works

Effective Attribute Quest!

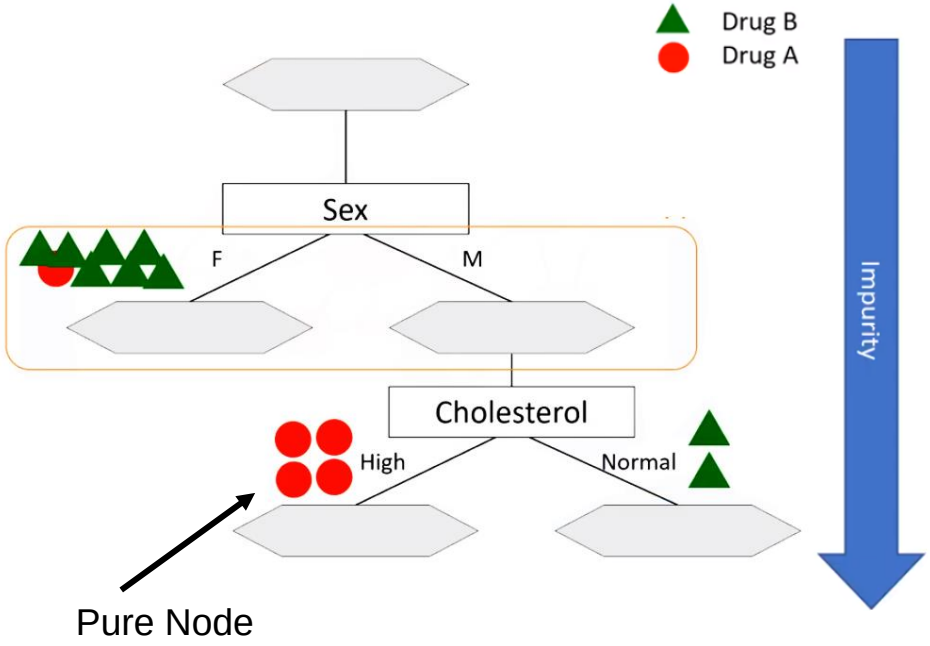
- Let’s try with “sex” attribute of patients.
- However, it is still a **better choice** in comparison with the “Cholesterol” attribute.
- **Reason:** as the result in the nodes are more pure, i.e. nodes which are either mostly Drug A or Drug B.
- Thus, it’s **more predictive** than the other attributes.



How Decision Tree Works

Effective Attribute Quest!

- Predictiveness is based on decrease in “impurity” of nodes.
- So, the Gender based feature is a good candidate in the following case, because it almost found the pure patients.
- We test “Cholesterol” again here.
- As you can see, it results in even more pure leaves.
- So, we can easily make a decision here.



How Decision Tree Works

Intuition of Node Impurity

- Method uses recursive partitioning to **minimize the "impurity"** at each step.
- "Impurity" of nodes is calculated by **"Entropy"** of data in the node.
- So, what is "Entropy"?
- Entropy is the amount of **information disorder**, or the amount of randomness in the data.
- In decision trees, we're looking for trees that have the smallest entropy in their nodes.
- **Lower the Entropy, distribution is less uniform and nodes are highly pure!**

Entropy

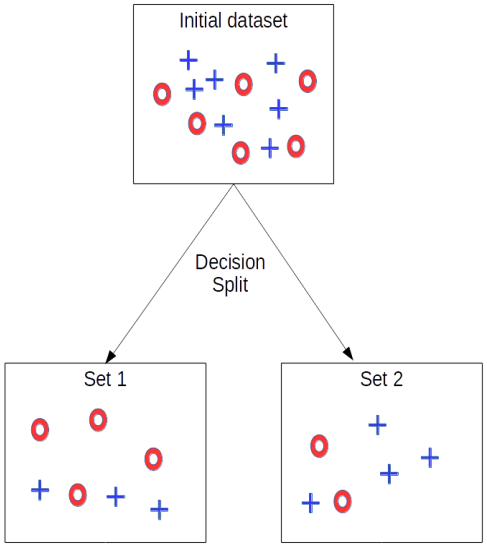
- To calculate entropy, formulae is:

$$Entropy = -p(A)\ \log(p(A)) - p(B)\ \log(p(B))\quad **base\ 2$$

- ***p*** is for the proportion or ratio of a category, such as Drug *A* or *B*.

Let's calculate the entropy of the dataset in our case, before splitting it.

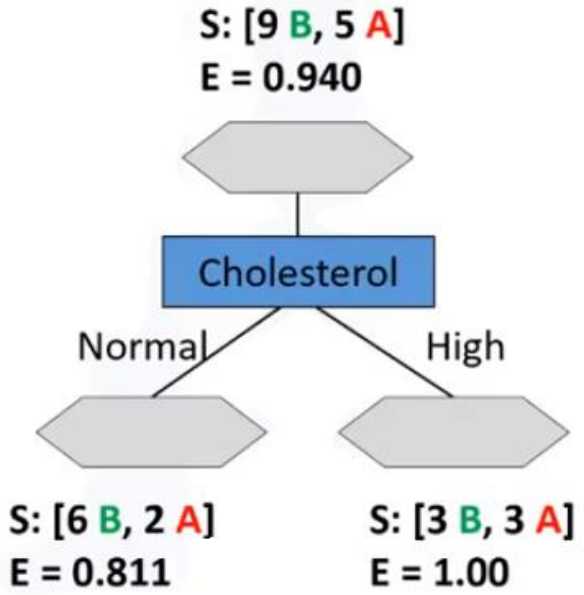
- We have 9 occurrences of Drug *B* and 5 of Drug *A*.
- Entropy = 0.530 + (0.410) = 0.940 (approx.).



Entropy in Decision Tree

Entropy at each Node (for each Attribute)

- Consider all the attributes and calculate the “Entropy” after the split, and then chose the best attribute.
- Calculate node entropy for Cholesterol feature.
- Separate test is conducted for each valid feature.



How Decision Tree Works

Entropy at Each Node (for each Attribute)

- Calculate node entropy for Sex feature

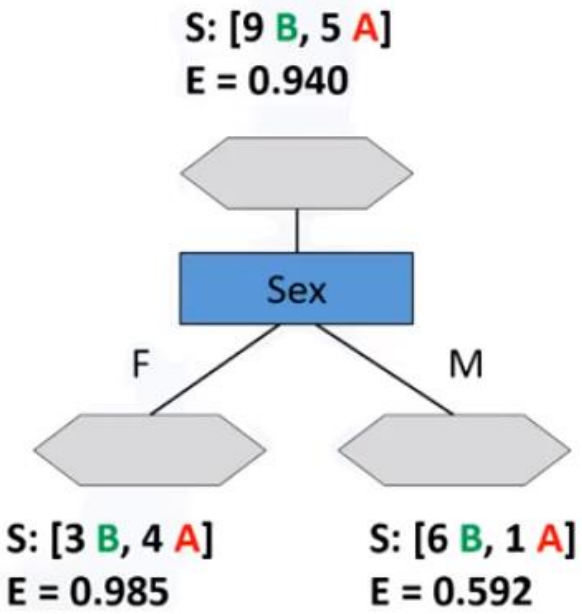
  -(9/14*log(9/14) + 5/14*log(5/14))

  =0.940

- Entropy of branch **F**

  -(3/7*log(3/7) + (4/7*log(4/7))

  =0.985



How Decision Tree Works

Information Gain (ID3)

- Before taking our splitting decision, lets understand **Information Gain**!
- Information gain is the information that can **increase the level of certainty** after splitting.
- As entropy, or the amount of randomness, decreases, the information gain, or amount of certainty, increases, and vice-versa.
- So, constructing a decision tree is all about finding attributes that return the highest information gain.
- **Information Gain = Total Entropy – Sum of Conditional Entropies**

Comparison of Attributes

Information Gain (Sex)

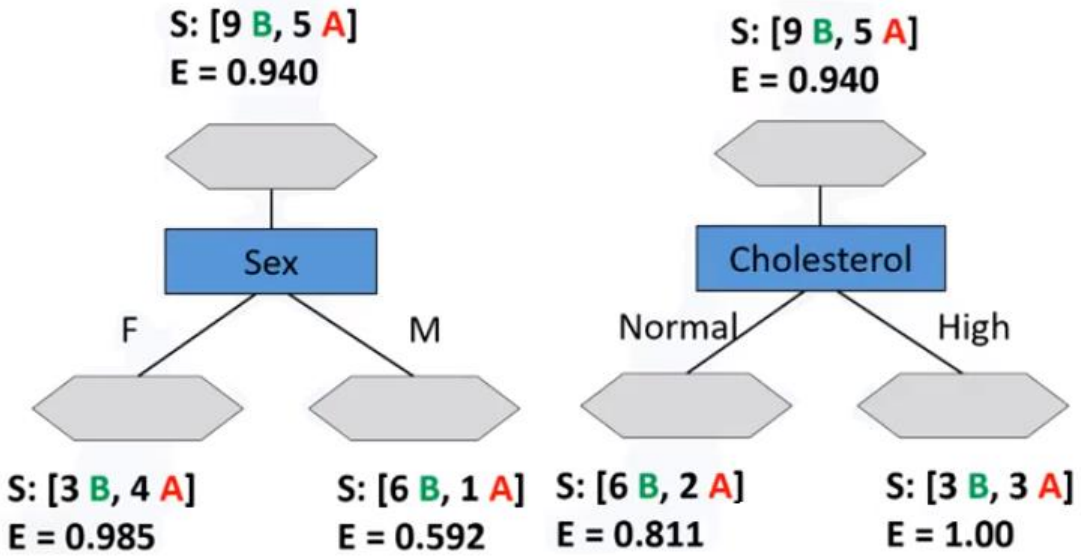
0.940-(7/14*0.985)-(7/14*0.592)

= **0.151**

Information Gain (Cholestrol)

0.940-(8/14*0.811)-(6/14*1)

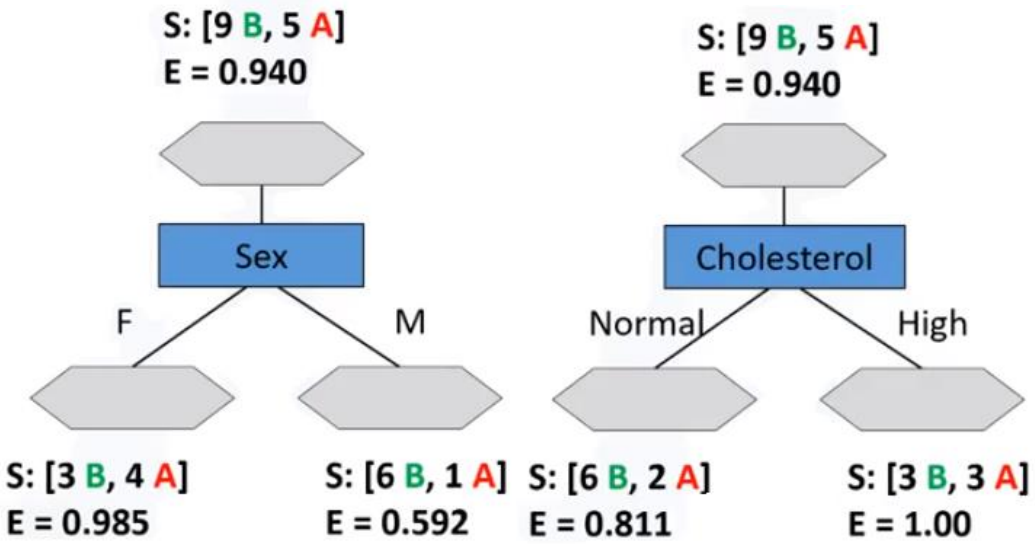
= **0.048**



How Decision Tree Works

Question ?

- Between the Cholesterol and Sex attributes, which one is a better choice?
- Which one is better as the first attribute to divide the dataset into 2 branches?
- Which attribute results in more pure nodes for our drugs?
- Answer: **“Sex”** attribute



How Decision Tree Works

Repeat!

- So, we select the “Sex” attribute as the first splitter.
- Now, what is the next attribute after branching by the “Sex” attribute?
- We should repeat the process for each branch, and test each of the other attributes to continue to reach the most pure leaves.
- This is the way that you build a decision tree!

Patient ID	Age	Sex	BP	Cholesterol	Drug
p1	Young	F	High	Normal	Drug A
p2	Young	F	High	High	Drug A
p3	Middle-age	F	Hiigh	Normal	Drug B
p4	Senior	F	Normal	Normal	Drug B
p5	Senior	M	Low	Normal	Drug B
p6	Senior	M	Low	High	Drug A
p7	Middle-age	M	Low	High	Drug B
p8	Young	F	Normal	Normal	Drug A
p9	Young	M	Low	Normal	Drug B
p10	Senior	M	Normal	Normal	Drug B
p11	Young	M	Normal	High	Drug B
p12	Middle-age	F	Normal	High	Drug B
p13	Middle-age	M	High	Normal	Drug B
p14	Senior	F	Normal	High	Drug A
p15	Middle-age	F	Low	Normal	?

Patient Drug Data

Lab 1 – Implement Decision Tree Machine Learning Algorithm

Summary

- Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems.
- It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules, and each leaf node represents the outcome.
- Entropy is the amount of information disorder or the amount of randomness in the data. The entropy in the node depends on how much random data is in that node and is calculated for each node.
- Information gain is the information that can **increase the level of certainty** after splitting.
- As entropy, or the amount of randomness, decreases, the information gain, or amount of certainty, increases, and vice-versa.

Quiz

1) **Decision trees are also known as CART. What is CART?**
   (A) Classification and Regression Trees
   (B) Customer Analysis and Research Tool
   (C) Communication Access Real-time Translation
   (D) Computerized Automatic Rating Technique

(A) Classification and Regression Trees

Quiz

2) Decision tree can be used for _____.
(A) classification
(B) regression
(C) Both
(D) None of these

C). Both

Quiz

3) Decision tree is a _____ algorithm.
(A) supervised learning
(B) unsupervised learning
(C) Both
(D) None of these

A). supervised learning

Quiz

4) Suppose, your target variable is whether a passenger will survive or not using Decision Tree. What type of tree do you need to predict the target variable?

(A) classification tree
(B) regression tree
(C) clustering tree
(D) dimensionality reduction tree

(A) Classification tree

Quiz

5) Suppose, your target variable is the price of a house using Decision Tree. What type of tree do you need to predict the target variable?

(A) classification tree
(B) regression tree
(C) clustering tree
(D) dimensionality reduction tree

(B) regression tree

Reference

https://kawsar34.medium.com/machine-learning-quiz-05-decision-tree-part-1-3ea71fa312e5

https://www.javatpoint.com

https://www.tutorialspoint.com

www.towardsdatascience.com

How Decision Tree Works !. In this Blog, I'll be covering the… | by Mehmet Toprak | Medium

Thank you...!